US010862928B1

(12) **United States Patent**
Badawy et al.

(10) **Patent No.: US 10,862,928 B1**
(45) **Date of Patent: Dec. 8, 2020**

(54) **SYSTEM AND METHOD FOR ROLE VALIDATION IN IDENTITY MANAGEMENT ARTIFICIAL INTELLIGENCE SYSTEMS USING ANALYSIS OF NETWORK IDENTITY GRAPHS**

(71) Applicant: **SAILPOINT TECHNOLOGIES, INC.**, Wilmington, DE (US)

(72) Inventors: **Mohamed M. Badawy**, Round Rock, TX (US); **Jostine Fei Ho**, Austin, TX (US); **Rajat Kabra**, Austin, TX (US)

(73) Assignee: **SAILPOINT TECHNOLOGIES, INC.**, Wilmington, DE (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/900,530**

(22) Filed: **Jun. 12, 2020**

(51) **Int. Cl.**

H04L 29/06	(2006.01)
G06N 5/04	(2006.01)
G06N 5/02	(2006.01)
G06F 21/60	(2013.01)
G06N 20/00	(2019.01)

(52) **U.S. Cl.**
CPC ***H04L 63/205*** (2013.01); ***G06F 21/604*** (2013.01); ***G06N 5/02*** (2013.01); ***G06N 5/04*** (2013.01); ***H04L 63/102*** (2013.01); ***H04L 63/104*** (2013.01); *G06N 20/00* (2019.01)

(58) **Field of Classification Search**
CPC ... H04L 63/102; H04L 63/104; H04L 63/205; G06N 5/02; G06N 5/04; G06N 20/00; G06F 21/604
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,526,785	B1	4/2009	Pearson et al.	
7,725,416	B2 *	5/2010	Buss	G06N 5/025 706/47
8,065,712	B1	11/2011	Cheng	
8,209,742	B2	6/2012	Schreiber et al.	
8,656,161	B2	2/2014	Nakae	
8,683,546	B2	3/2014	Dunagan	
9,077,728	B1	7/2015	Hart	
9,246,945	B2	1/2016	Chari et al.	
9,286,595	B2 *	3/2016	Taneja	G06Q 10/103
9,288,232	B2	3/2016	Chari et al.	
9,516,053	B1 *	12/2016	Muddu	G06F 16/254
9,679,125	B2	6/2017	Bailor	
9,699,196	B1	7/2017	Kolman	

(Continued)

OTHER PUBLICATIONS

Saltzer, Jerome H. et al., "The Protection of Information in Computer Systems," Proceedings of the IEEE, 63(9), Sep. 1975, pp. 1278-1308.

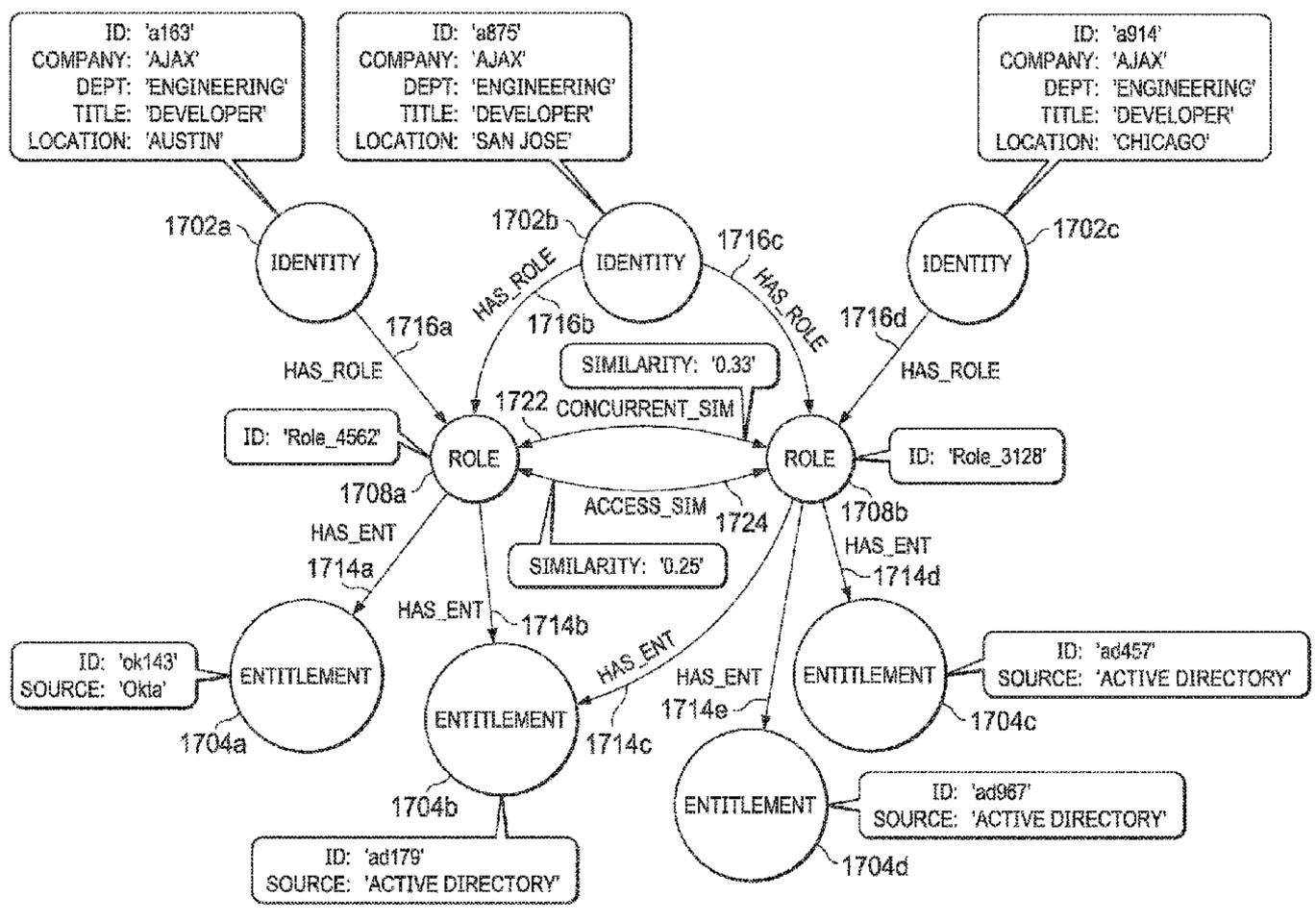
(Continued)

Primary Examiner — Linglan E Edwards
(74) *Attorney, Agent, or Firm* — Sprinkle IP Law Group

(57) **ABSTRACT**

Systems and methods for embodiments of a graph based artificial intelligence systems for identity management are disclosed. Embodiments of the identity management systems disclosed herein may utilize a network graph approach to analyzing roles of a distributed networked enterprise computing environment. Specifically, in certain embodiments, an artificial intelligence based identity management systems may utilize role graphs to assess the role structure of a distributed enterprise computing environment.

21 Claims, 26 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

9,787,688	B2 *	10/2017	Jagtap	G06F 21/6218
9,819,685	B1	11/2017	Scott	
9,992,230	B1	6/2018	Haverty	
10,298,589	B2 *	5/2019	Cleaver	H04L 63/104
1,034,143	A1	7/2019	Badawy et al.	
10,341,430	B1 *	7/2019	Badawy	G06F 16/906
10,348,735	B2 *	7/2019	Anderson	H04L 63/06
10,476,952	B1 *	11/2019	Badawy	G06F 21/45
10,476,953	B1 *	11/2019	Badawy	G06F 21/45
10,523,682	B1 *	12/2019	Badawy	G06F 21/316
10,554,665	B1 *	2/2020	Badawy	G06F 16/9024
10,621,368	B2 *	4/2020	Ravizza	G06F 16/9024
10,681,056	B1	6/2020	Badawy	
10,791,170	B2	9/2020	Badawy	
2002/0026592	A1	2/2002	Gavrila et al.	
2007/0067845	A1	3/2007	Wiemer	
2007/0226248	A1	9/2007	Darr	
2008/0091681	A1	4/2008	Dwivedi et al.	
2008/0147584	A1 *	6/2008	Buss	G06N 5/025 706/47
2008/0288330	A1	11/2008	Hildebrand	
2009/0222894	A1	9/2009	Kenny et al.	
2009/0300711	A1	12/2009	Tokutani et al.	
2010/0082695	A1	4/2010	Hardt	
2010/0274815	A1	10/2010	Vanasco	
2011/0209196	A1	8/2011	Kennedy	
2012/0023576	A1	1/2012	Sorensen et al.	
2012/0216243	A1	8/2012	Gill et al.	
2012/0246098	A1	9/2012	Chari et al.	
2013/0232539	A1	9/2013	Polunin	
2013/0254833	A1	9/2013	Nicodemus et al.	
2013/0283339	A1	10/2013	Biswas et al.	
2014/0207813	A1	7/2014	Long	
2015/0128211	A1	5/2015	Kirner	
2015/0379429	A1	12/2015	Lee	
2016/0203327	A1	7/2016	Akkiraju	
2016/0294645	A1	10/2016	Kirner	
2016/0294646	A1	10/2016	Kirner	
2017/0103164	A1 *	4/2017	Dunlevy	G06F 16/24573
2017/0103165	A1 *	4/2017	Dunlevy	G06F 16/2365
2017/0147790	A1 *	5/2017	Patel	H04L 63/105
2017/0220964	A1	8/2017	Datta Ray	
2017/0310552	A1	10/2017	Wallerstein	
2017/0329957	A1 *	11/2017	Vepa	G06F 21/34
2017/0364534	A1	12/2017	Zhang	
2018/0069899	A1	3/2018	Lang	
2019/0114342	A1	4/2019	Orun	
2020/0007555	A1 *	1/2020	Jadhav	H04L 67/10
2020/0169565	A1 *	5/2020	Badawy	H04L 63/104
2020/0169603	A1	5/2020	Badawy	
2020/0259840	A1	8/2020	Badawy	
2020/0274880	A1	8/2020	Badawy	
2020/0280565	A1	9/2020	Badawy	

OTHER PUBLICATIONS

Bishop, Matt et al., "We have Met the Enemy and He is Us," NSPW '08: Proceedings of the 2008 workshop on New Security paradigms, Sep. 2008, 11 pgs.

Frank, Mario et al., "A probabilistic approach to hybrid role mining," CCS '09, Nov. 2009, 11 pgs.

Molloy, Ian et al., "Generative Models for Access Control Policies: Applications to Role Mining Over Logs with Attribution," Proceedings of the 17th ACM Symposium on Access Control Models and Technologies, SACMAT, Jun. 2012, 11 pgs.

Blei, David M. et al., "Latent Dirichlet Allocation," Journal of Machine Learning Research 3, Jan. 2003, pp. 993-1022.

McDaniel, Patrick et al., "Securing Distributed Applications Using a Policy-based Approach," Ann Arbor, Dec. 19, 2003, 48109-2122, 24 pgs.

Chen, Ying et al., "Data Mining and Service Rating in Service-Oriented Architectures to Improve Information Sharing," 2005 IEEE Aerospace Conference, (Version 7, Updated Jan. 27, 2005) Mar. 2005, 11 pgs.

Molloy, Ian, "Automatic Migration to Role-Based Access Control," CERIAS Tech Report 2010-34, Purdue University, IN, Thesis Dissertation/Acceptance, Aug. 2010, 178 pgs.

Ene, Alina et al., "Fast Exact and Heuristic Methods for Role Minimization Problems," SACMAT '08 Proceedings of the 13th ACM symposium on Access control models and technologies, Estes, CO, Jun. 11-13, 2008, pp. 1-10.

Harrison, Michael A. et al., "Protection in Operating Systems," Communications of the ACM, vol. 19, No. 8, Aug. 1976, pp. 461-471.

Li, Ninghui et al., "Access Control Policy Combining: Theory Meets Practice," Proceedings of the 14th ACM symposium on Access control models and technologies SACMAT '09, Jun. 3-5, 2009, 10 pgs.

Schneider, Fred B., "Least Privilege and More," IEEE Security & Privacy, vol. 1, Issue 5, Sep. 2003, pp. 209-213.

Office Action for U.S. Appl. No. 13/970,174, dated Nov. 4, 2014, 22 pgs.

Office Action for U.S. Appl. No. 13/904,350, dated Nov. 14, 2014, 33 pgs.

Office Action for U.S. Appl. No. 13/904,350, dated Apr. 27, 2015, 28 pgs.

Office Action for U.S. Appl. No. 13/970,174, dated Jun. 24, 2015, 16 pgs.

Xu, Wei et al., Taint-Enhanced Policy Enforcement: A Practical Approach to Defeat a Wide Range of Attacks, 15th USENIX Security Symposium, 2006, pp. 121-136.

International Search Report and Written Opinion for International Patent Application No. PCT/US2019/062743, dated Jan. 13, 2020, 6 pgs.

Notice of Allowance for U.S. Appl. No. 16/582,493, dated Jun. 24, 2020, 4 pgs.

Notice of Allowance for U.S. Appl. No. 16/714,435, dated Jul. 8, 2020, 10 pgs.

* cited by examiner

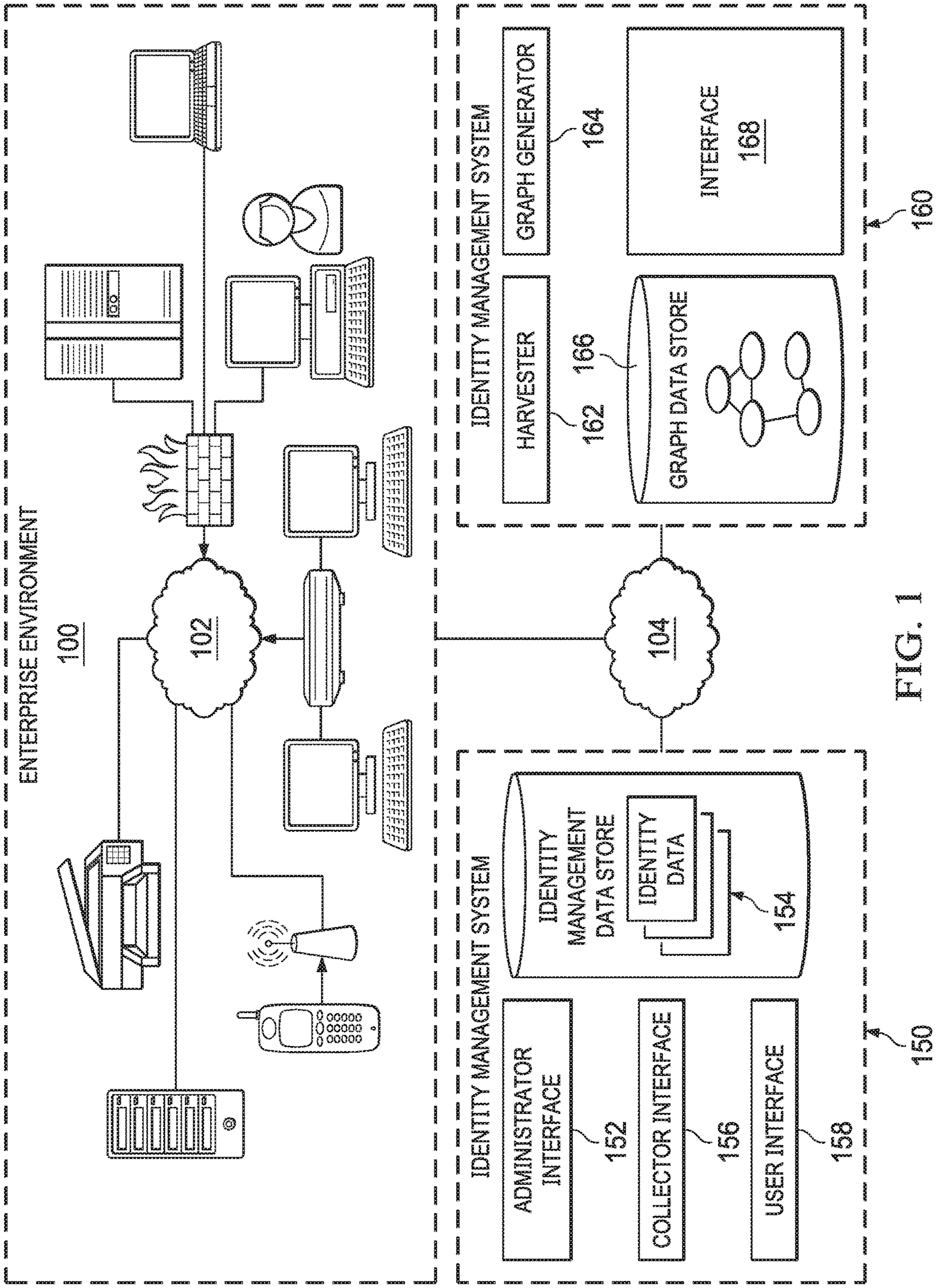

ENTERPRISE ENVIRONMENT
100
102
104
IDENTITY MANAGEMENT SYSTEM
GRAPH GENERATOR
164
INTERFACE
168
160
HARVESTER
162
166
GRAPH DATA STORE
IDENTITY MANAGEMENT DATA STORE
IDENTITY DATA
154
ADMINISTRATOR INTERFACE
152
COLLECTOR INTERFACE
156
USER INTERFACE
158
150

FIG. 1

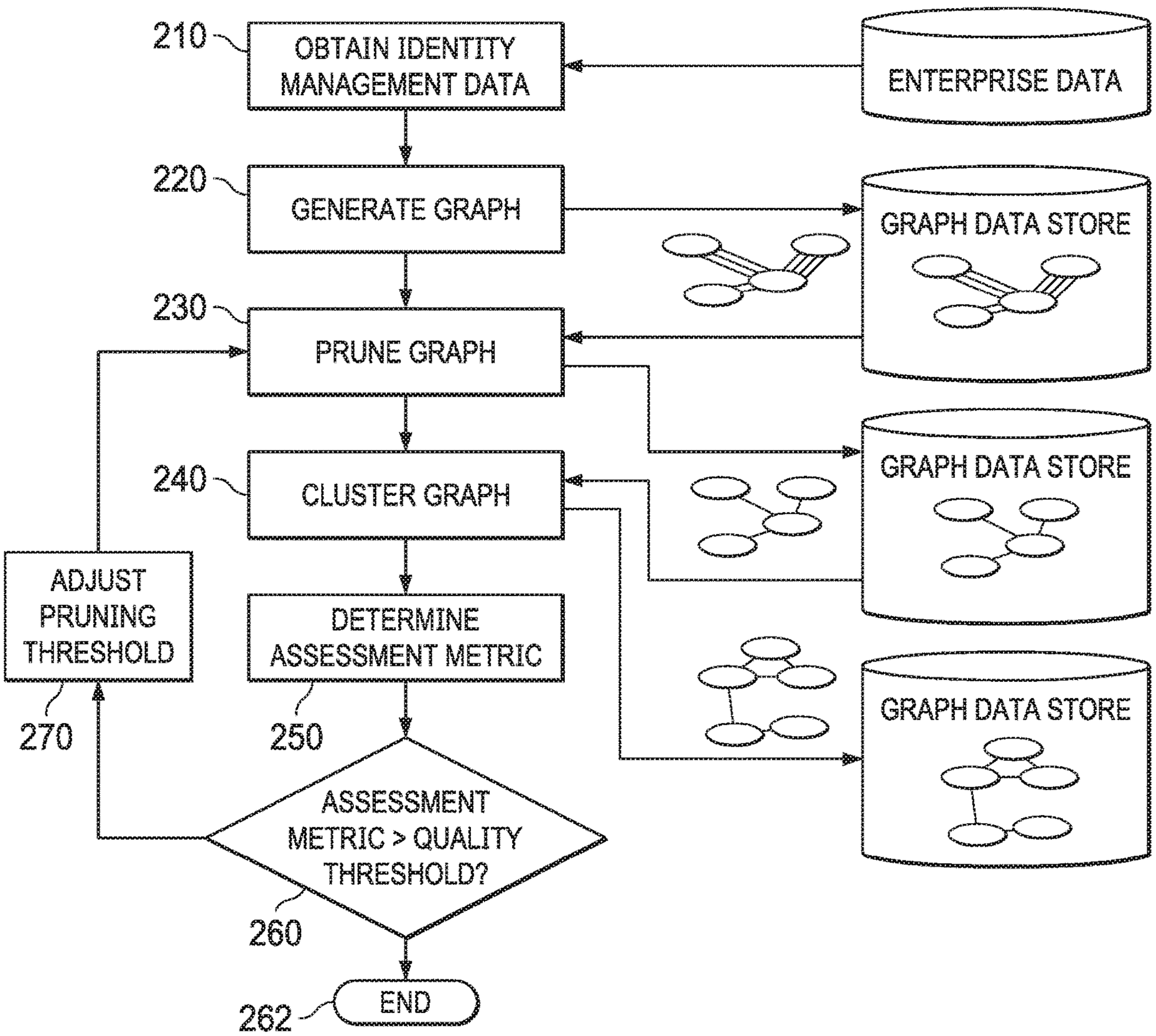

210
OBTAIN IDENTITY MANAGEMENT DATA
ENTERPRISE DATA
220
GENERATE GRAPH
GRAPH DATA STORE
230
PRUNE GRAPH
240
CLUSTER GRAPH
GRAPH DATA STORE
ADJUST PRUNING THRESHOLD
270
DETERMINE ASSESSMENT METRIC
250
GRAPH DATA STORE
ASSESSMENT METRIC > QUALITY THRESHOLD?
260
END
262

FIG. 2

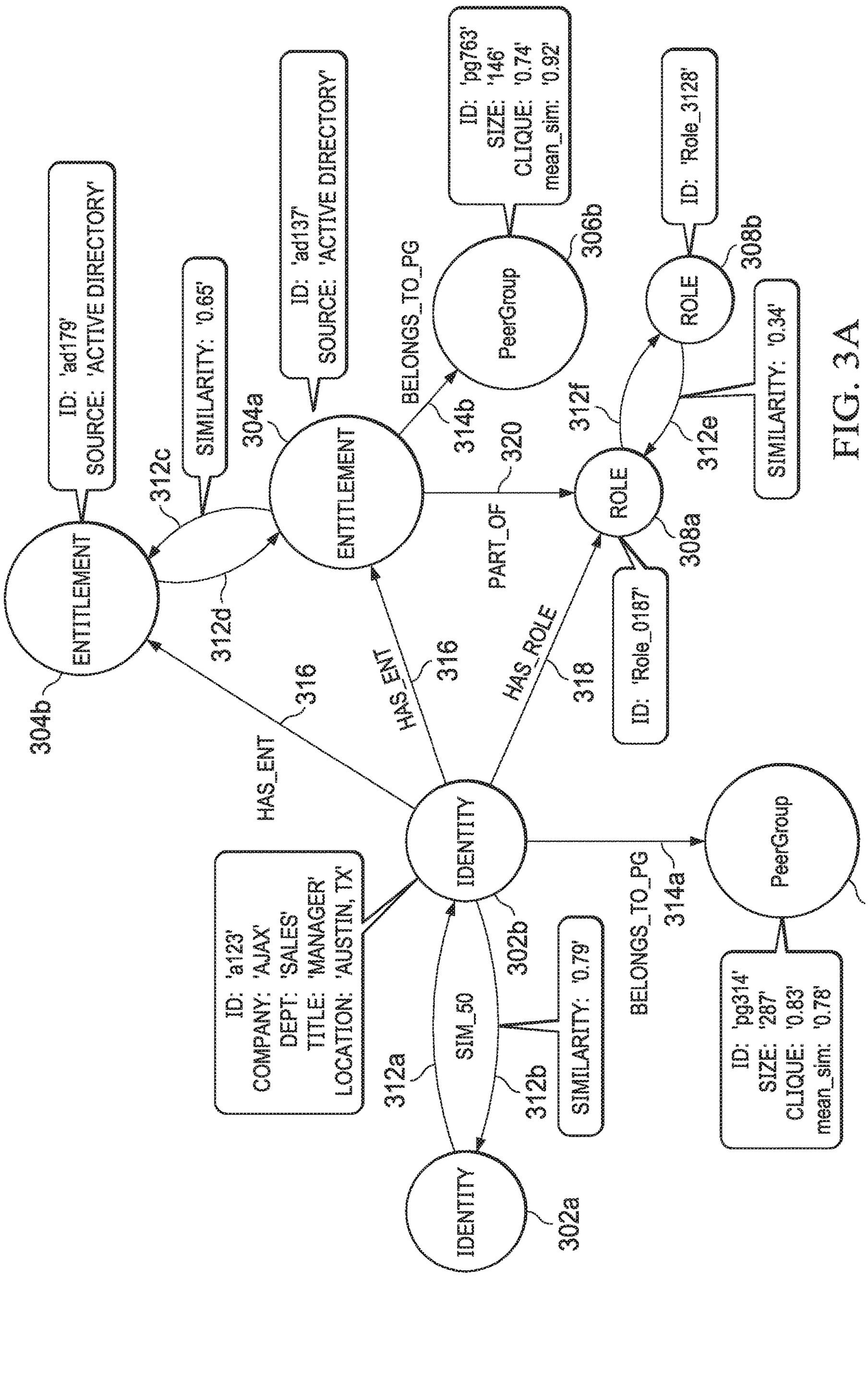

ID: 'ad179'
SOURCE: 'ACTIVE DIRECTORY'
SIMILARITY: '0.65'
ID: 'ad137'
SOURCE: 'ACTIVE DIRECTORY'
ID: 'pg763'
SIZE: '146'
CLIQUE: '0.74'
mean_sim: '0.92'
ID: 'Role_3128'
ID: 'a123'
COMPANY: 'AJAX'
DEPT: 'SALES'
TITLE: 'MANAGER'
LOCATION: 'AUSTIN, TX'
SIMILARITY: '0.79'
SIMILARITY: '0.34'
ID: 'Role_0187'
ID: 'pg314'
SIZE: '287'
CLIQUE: '0.83'
mean_sim: '0.78'
ENTITLEMENT
ENTITLEMENT
ENTITLEMENT
IDENTITY
IDENTITY
ROLE
ROLE
ROLE
PeerGroup
PeerGroup
HAS_ENT
HAS_ENT
HAS_ENT
HAS_ROLE
PART_OF
BELONGS_TO_PG
BELONGS_TO_PG
SIM_50
304a
304b
306a
306b
308a
308b
312a
312b
312c
312d
312e
312f
314a
314b
316
316
318
320
302a
302b

FIG. 3A

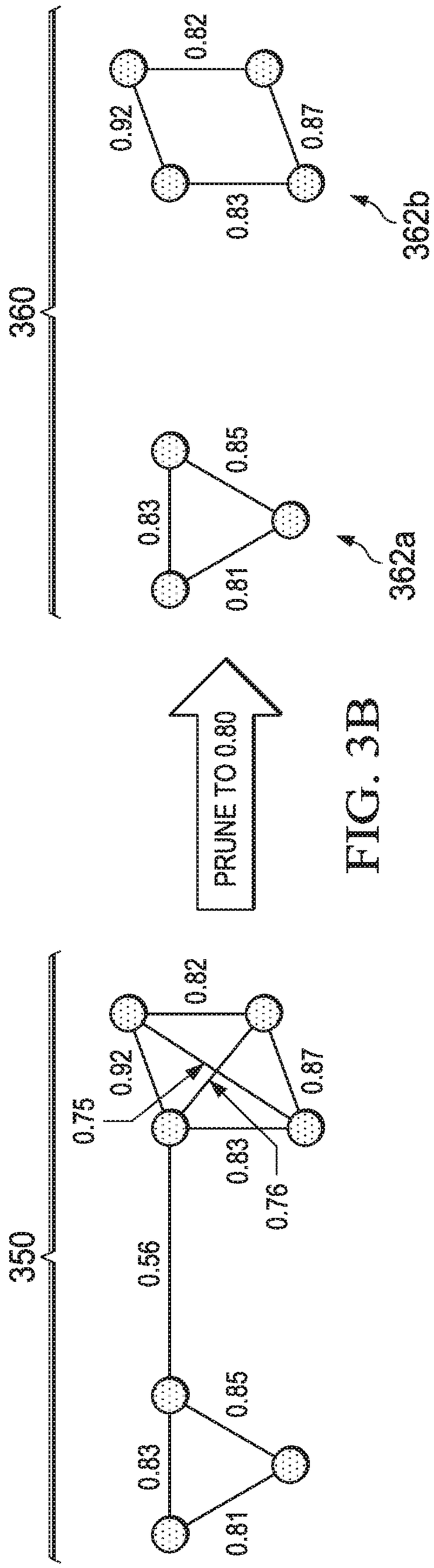

360
0.82
0.92
0.87
0.83
362b
0.85
0.83
0.81
362a
PRUNE TO 0.80
350
0.82
0.75
0.92
0.87
0.83
0.76
0.56
0.85
0.83
0.81

FIG. 3B

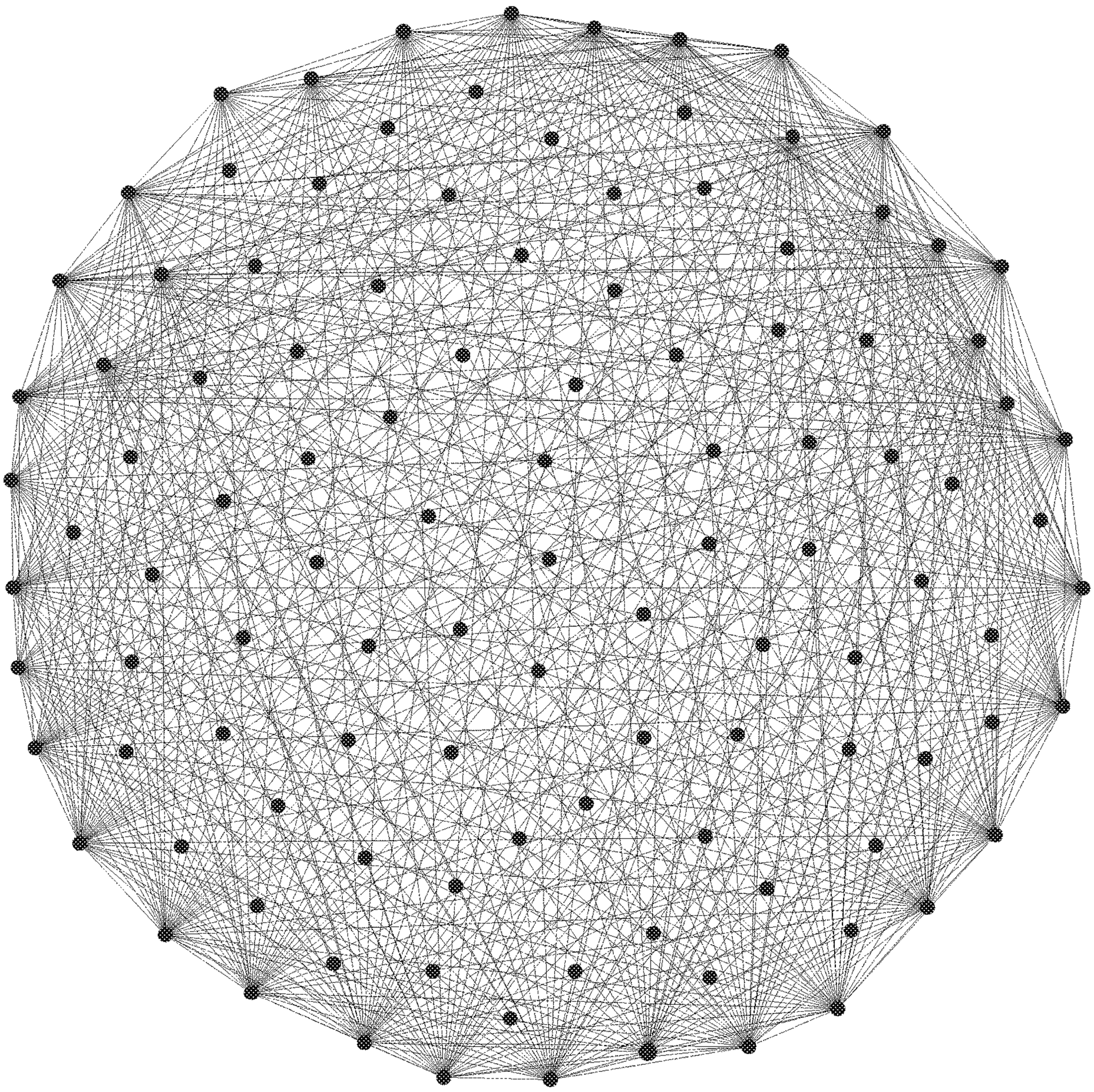


FIG. 3C

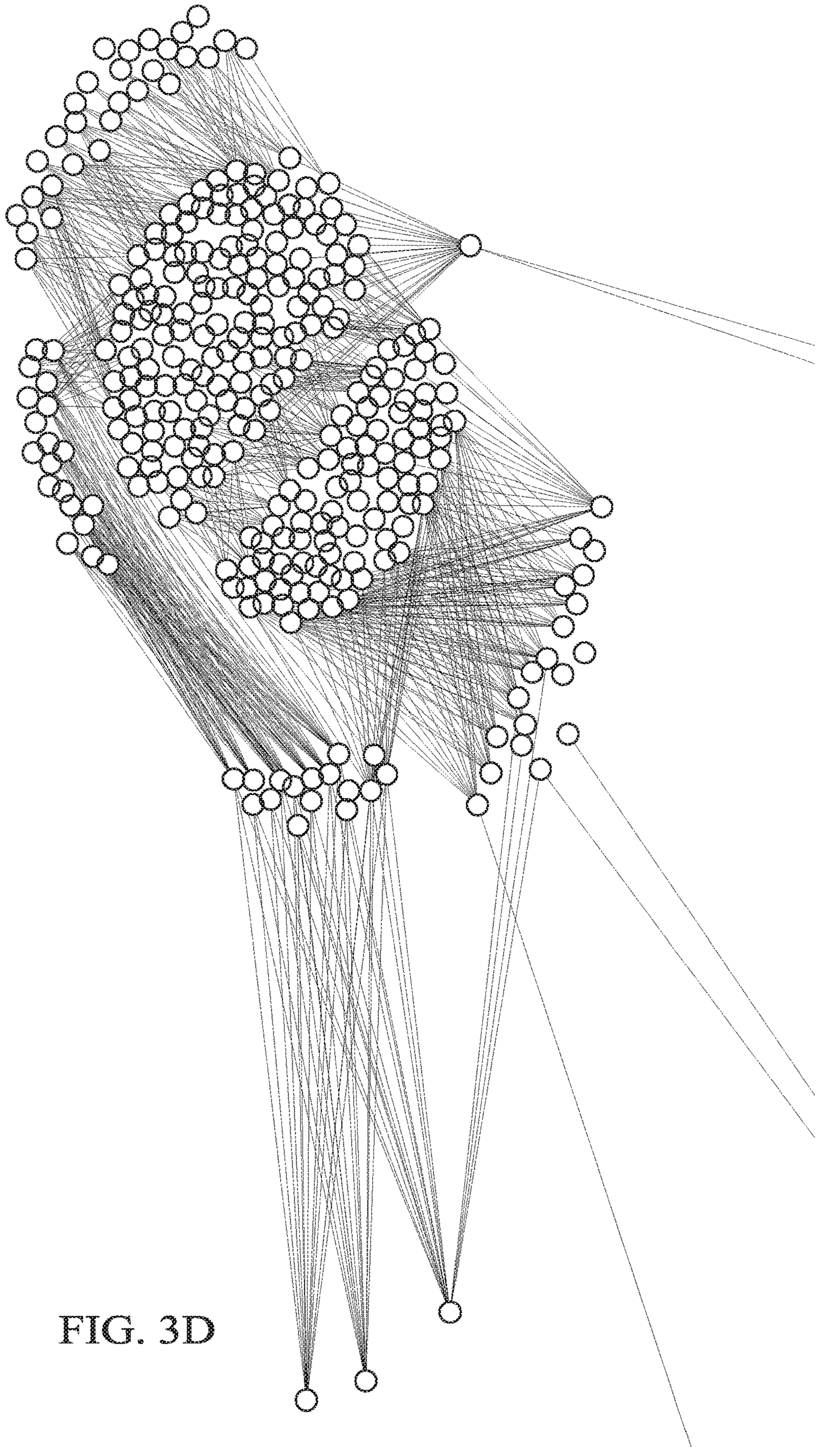


FIG. 3D

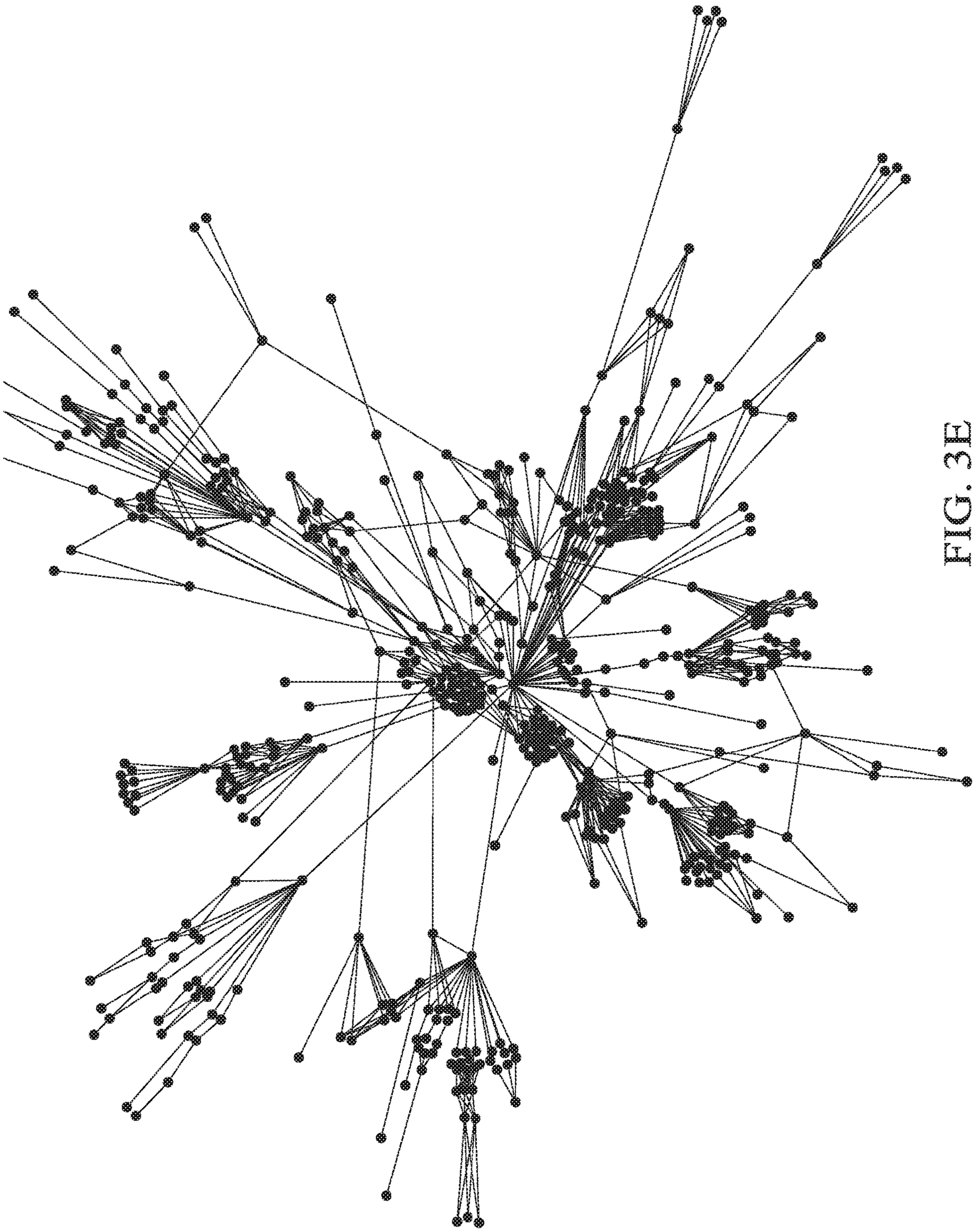


FIG. 3E

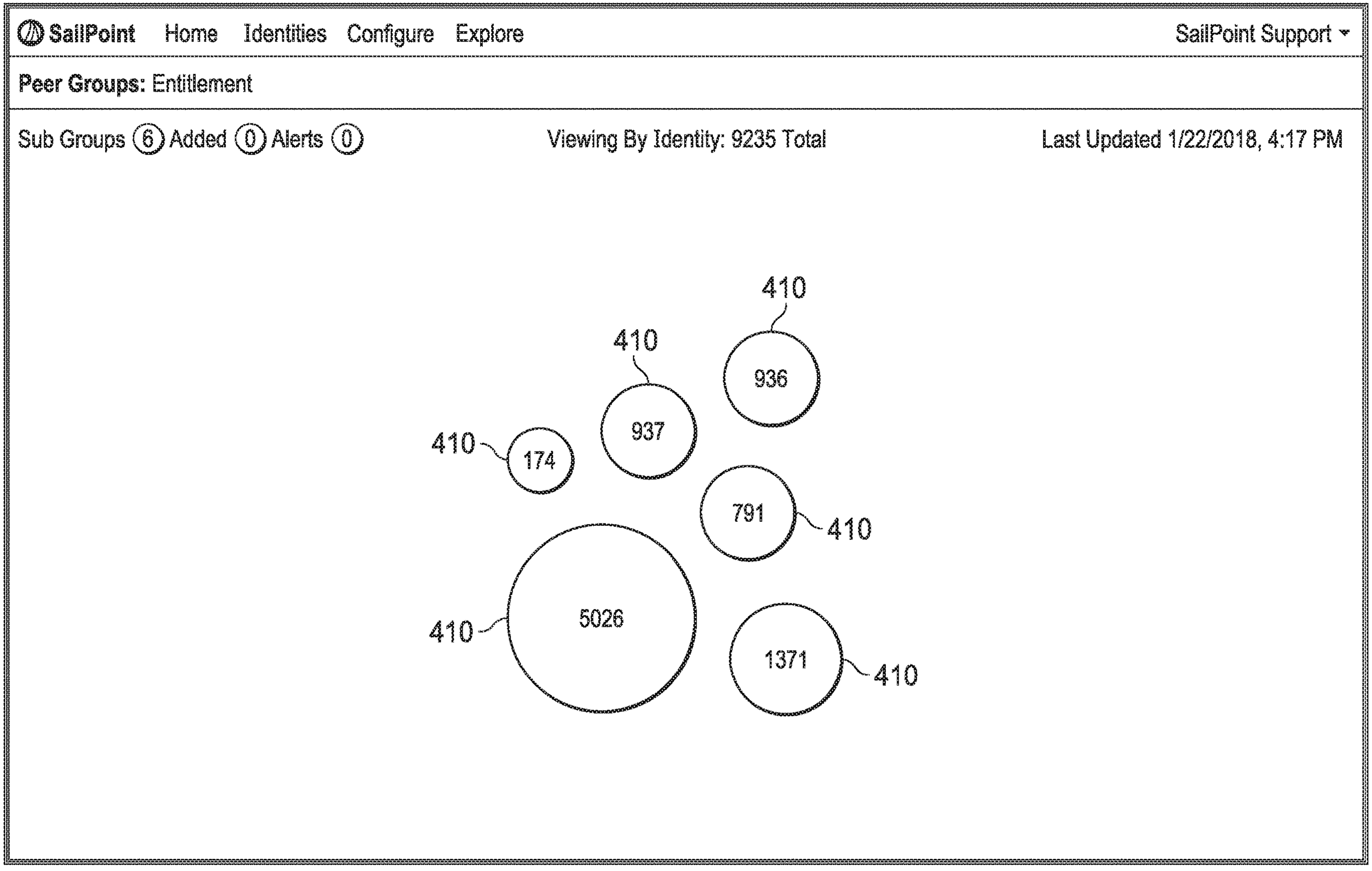

SailPoint
Home
Identities
Configure
Explore
SailPoint Support
Peer Groups: Entitlement
Sub Groups 6 Added 0 Alerts 0
Viewing By Identity: 9235 Total
Last Updated 1/22/2018, 4:17 PM
410
936
410
937
410
174
791
410
410
5026
1371
410

FIG. 4

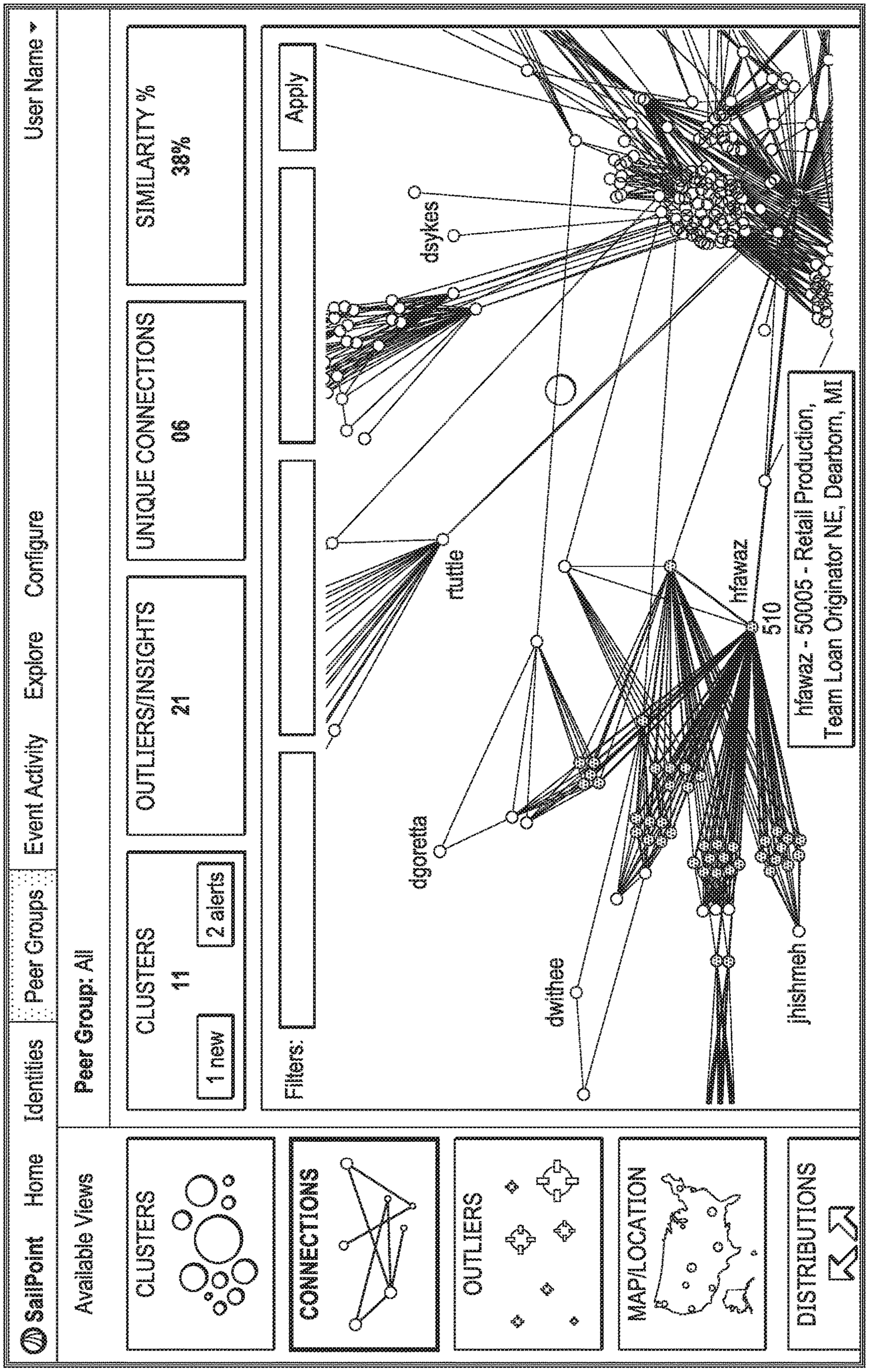

SailPoint
Home
Identities
Peer Groups
Event Activity
Explore
Configure
Available Views
CLUSTERS
CONNECTIONS
OUTLIERS
MAP/LOCATION
DISTRIBUTIONS
Peer Group: All
CLUSTERS
11
1 new
2 alerts
OUTLIERS/INSIGHTS
21
UNIQUE CONNECTIONS
06
SIMILARITY %
38%
Apply
Filters:
User Name
dsykes
rtuttle
dgoretta
hfawaz
510
dwithee
jhishmeh
hfawaz - 50005 - Retail Production,
Team Loan Originator NE, Dearborn, MI

FIG. 5

SailPoint Home Identities Configure Explore SailPoint Support

Peer Group: 937

Total Identities	Total Entitlements	Most Common Entitlement	Identities with Unique Entitlements
937	3577	Office365/PowerPoint access (from Office365)	128

Top 5 Representative Entitlements (Peer Group / All Identities)

	Peer Group	All Identities
Active Directory/Consumer Credit access (from Active Directory)	89%	56%
Active Directory/Treasury access (from Active Directory)	89%	56%
Active Directory/Financial Planning access (from Active Directory)	88%	56%
Active Directory/Loan Processing access (from Active Directory)	88%	56%
Active Directory/Commercial Loans access (from Active Directory)	88%	56%

Top 5 Most Common Entitlements (Peer Group / All Identities)

	Peer Group	All Identities
Office365/PowerPoint access (from Office365)	97%	91%
Office365/Word access (from Office365)	97%	91%
Office365/Email access (from Office365)	96%	91%
Office365/Excel access (from Office365)	96%	91%
Office365/Scheduler access (from Office365)	95%	91%

Least Frequently Assigned	Most Frequently Assigned	Entitlement Sources
Teller Front/Geneva Account creation access (from Teller Front) (1)	Office365/PowerPoint access (from Office365) (908)	

FIG. 6

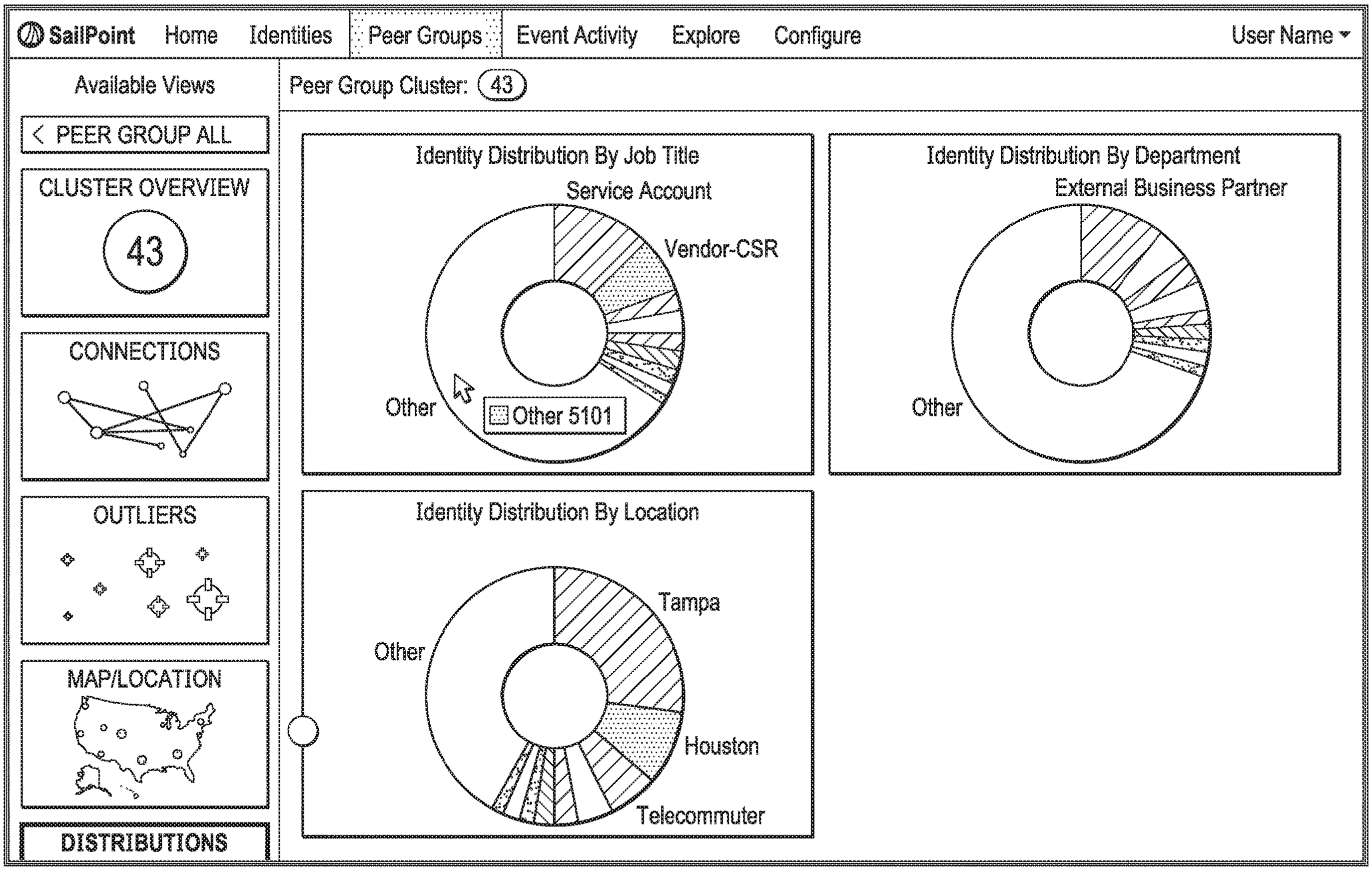

SailPoint
Home
Identities
Peer Groups
Event Activity
Explore
Configure
User Name
Available Views
Peer Group Cluster: 43
< PEER GROUP ALL
CLUSTER OVERVIEW
43
CONNECTIONS
OUTLIERS
MAP/LOCATION
DISTRIBUTIONS
Identity Distribution By Job Title
Service Account
Vendor-CSR
Other
Other 5101
Identity Distribution By Department
External Business Partner
Other
Identity Distribution By Location
Tampa
Other
Houston
Telecommuter

FIG. 7

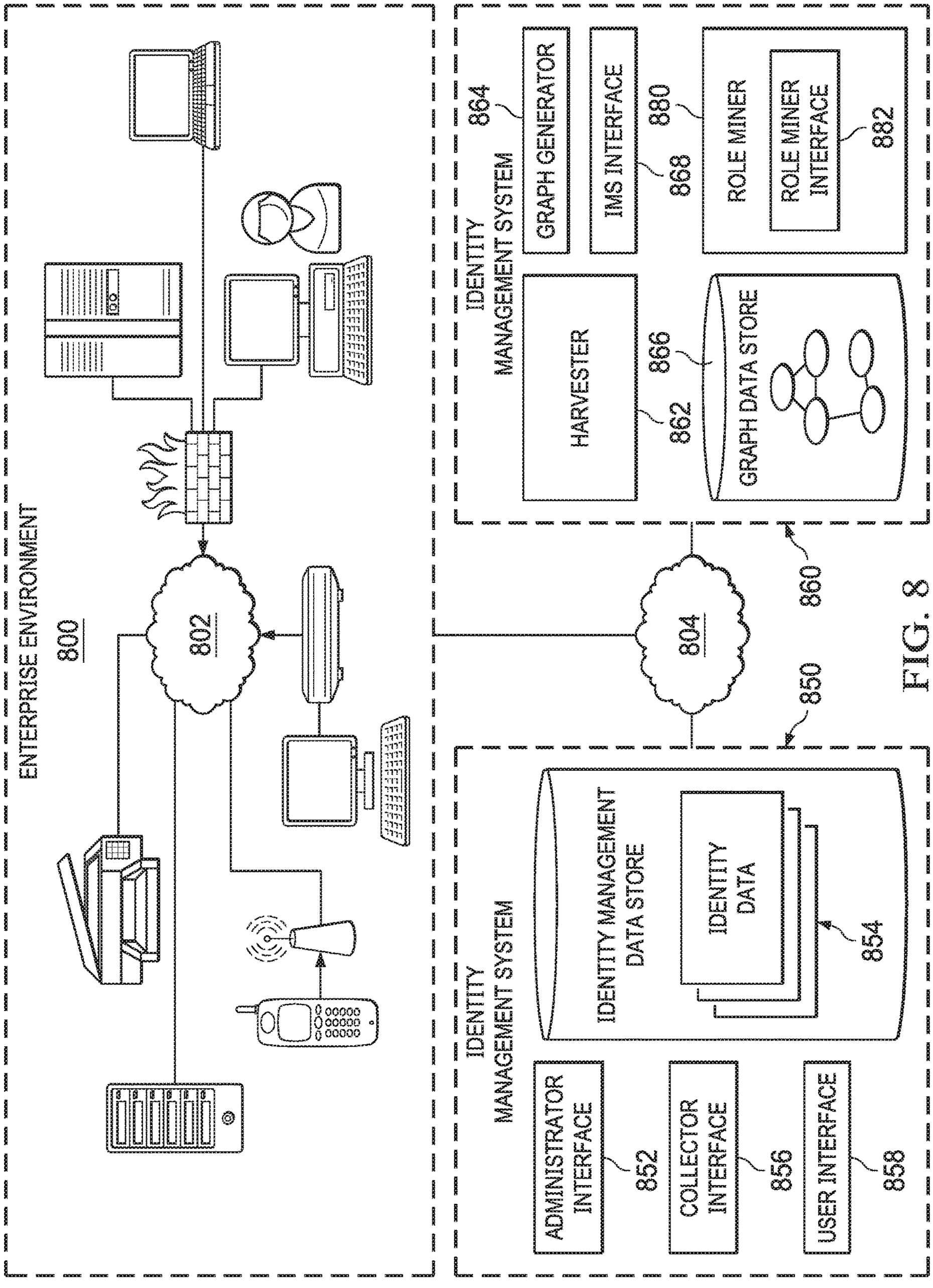

ENTERPRISE ENVIRONMENT
800
802
804
IDENTITY MANAGEMENT SYSTEM
GRAPH GENERATOR
864
IMS INTERFACE
880
868
ROLE MINER
ROLE MINER INTERFACE
882
HARVESTER
862
866
GRAPH DATA STORE
860
850
IDENTITY MANAGEMENT DATA STORE
IDENTITY DATA
854
ADMINISTRATOR INTERFACE
852
COLLECTOR INTERFACE
856
USER INTERFACE
858

FIG. 8

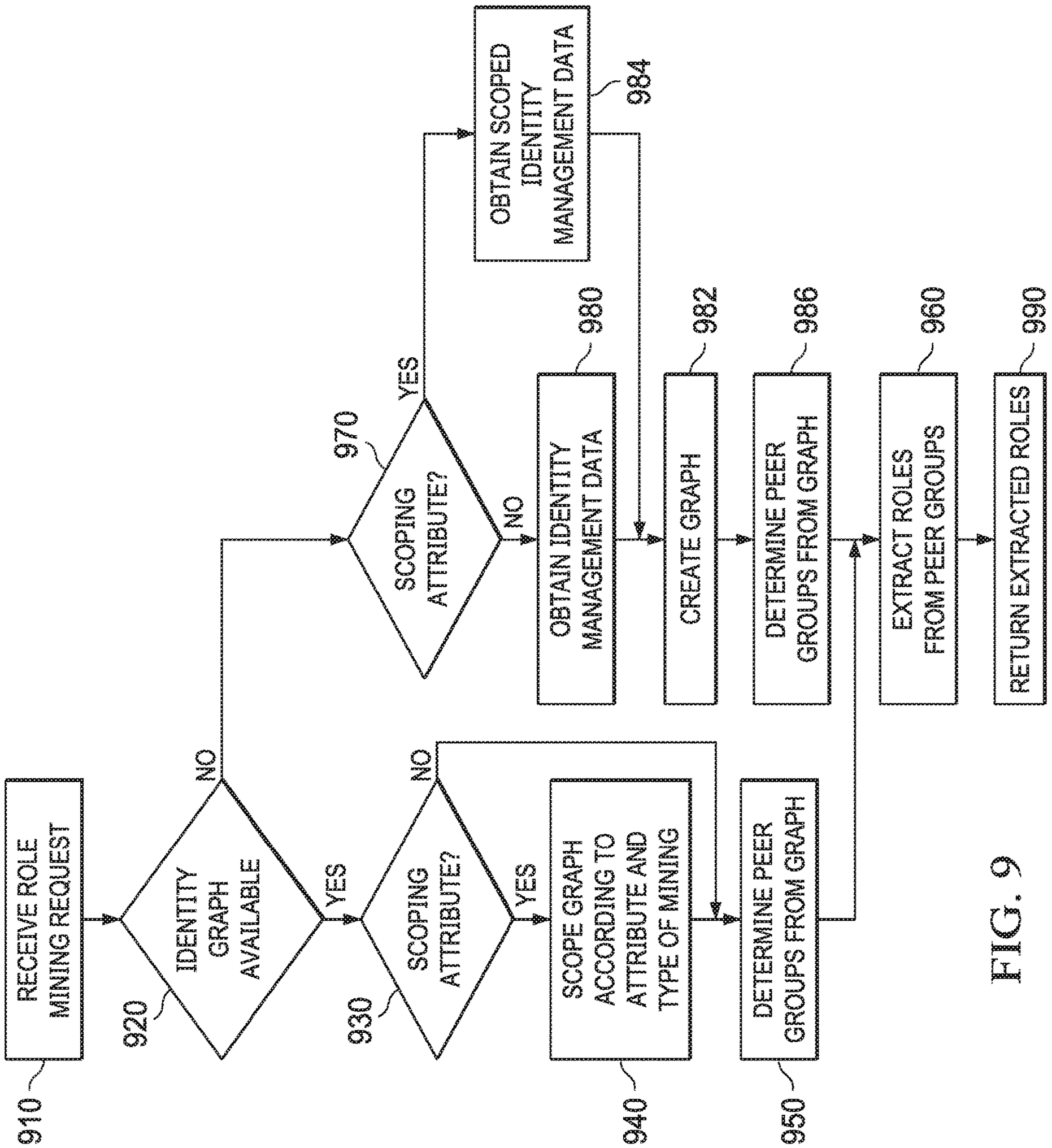

RECEIVE ROLE MINING REQUEST
910
IDENTITY GRAPH AVAILABLE
920
NO
YES
SCOPING ATTRIBUTE?
930
NO
YES
SCOPE GRAPH ACCORDING TO ATTRIBUTE AND TYPE OF MINING
940
DETERMINE PEER GROUPS FROM GRAPH
950
EXTRACT ROLES FROM PEER GROUPS
960
RETURN EXTRACTED ROLES
990
SCOPING ATTRIBUTE?
970
NO
YES
OBTAIN IDENTITY MANAGEMENT DATA
980
CREATE GRAPH
982
DETERMINE PEER GROUPS FROM GRAPH
986
OBTAIN SCOPED IDENTITY MANAGEMENT DATA
984

FIG. 9

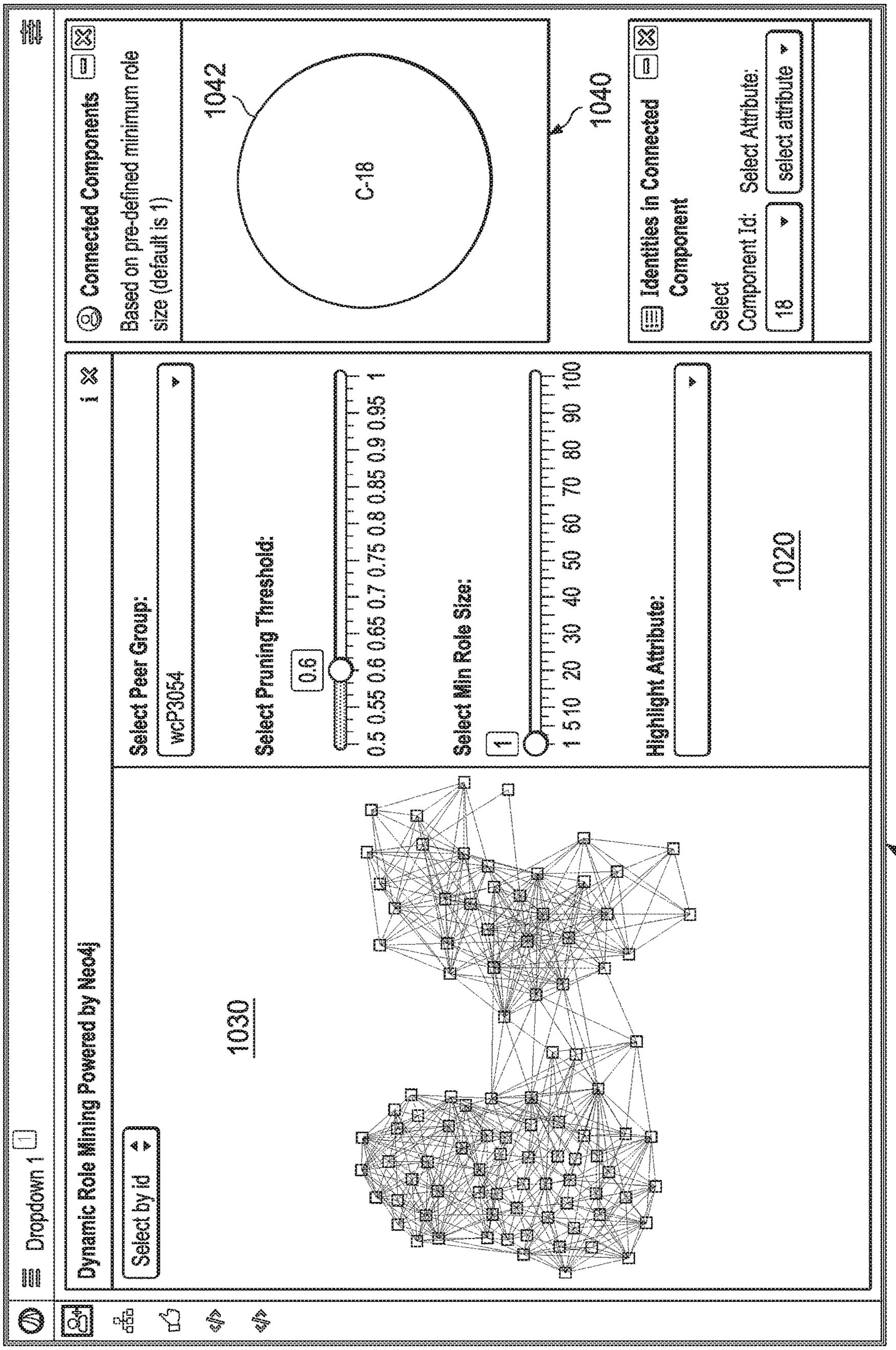

Connected Components
Based on pre-defined minimum role size (default is 1)
1042
C-18
1040
Identities in Connected Component
Select Component Id:
18
Select Attribute:
select attribute
Neo4j
Dropdown 1
Dynamic Role Mining Powered by Neo4j
Select by id
1030
Select Peer Group:
wcP3054
Select Pruning Threshold:
0.6
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Select Min Role Size:
1
1 5 10 20 30 40 50 60 70 80 90 100
Highlight Attribute:
1020
1010

FIG. 10

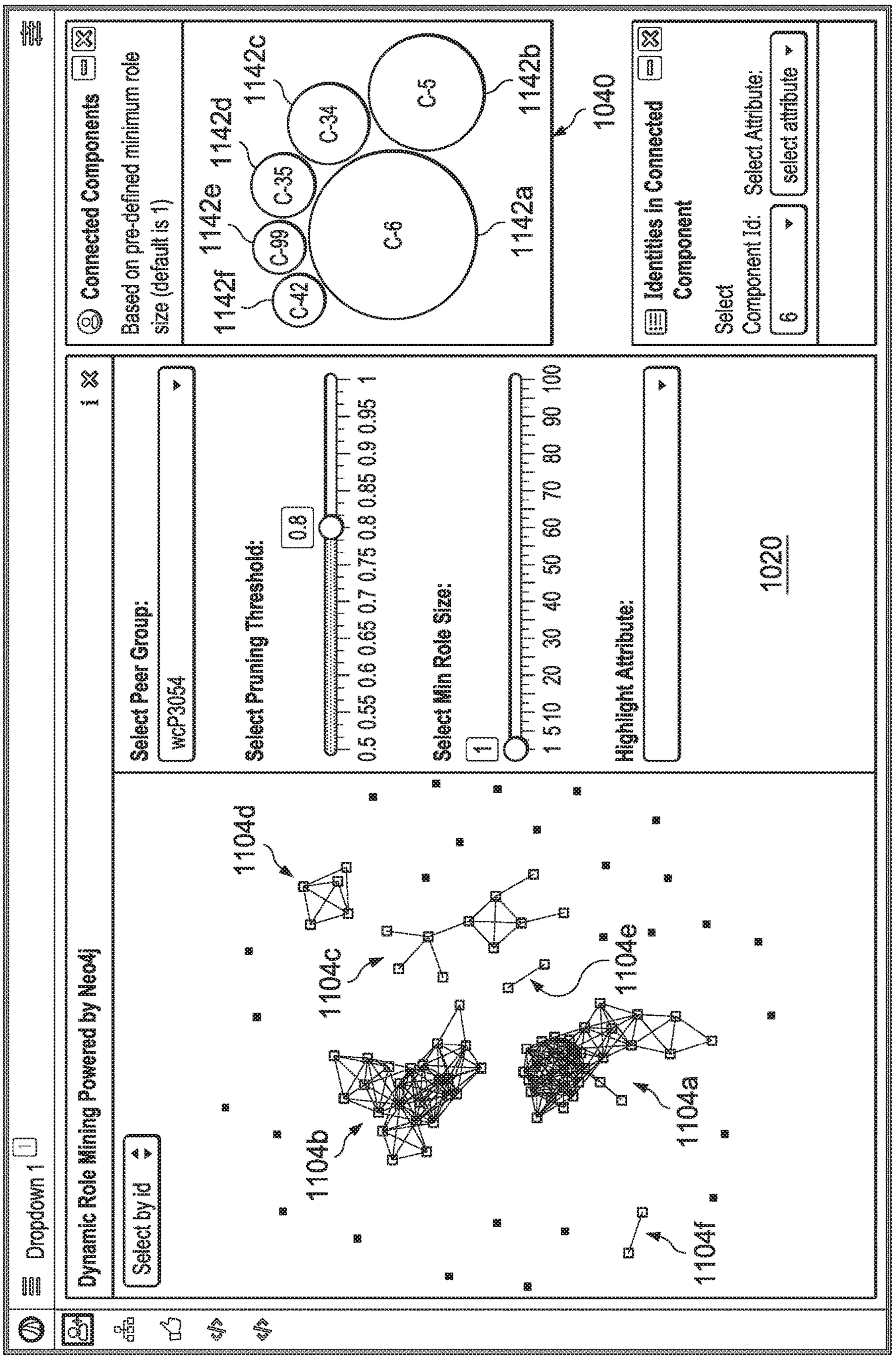

Connected Components
Based on pre-defined minimum role size (default is 1)
1142f
1142e
1142d
1142c
1142b
1142a
C-42
C-99
C-35
C-34
C-5
C-6
1040
Identities in Connected Component
Select Component Id:
6
Select Attribute:
select attribute
Select Peer Group:
wcP3054
Select Pruning Threshold:
0.8
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Select Min Role Size:
1
1 5 10 20 30 40 50 60 70 80 90 100
Highlight Attribute:
1020
Dropdown 1
Dynamic Role Mining Powered by Neo4j
Select by id
1104d
1104c
1104e
1104b
1104a
1104f

FIG. 11

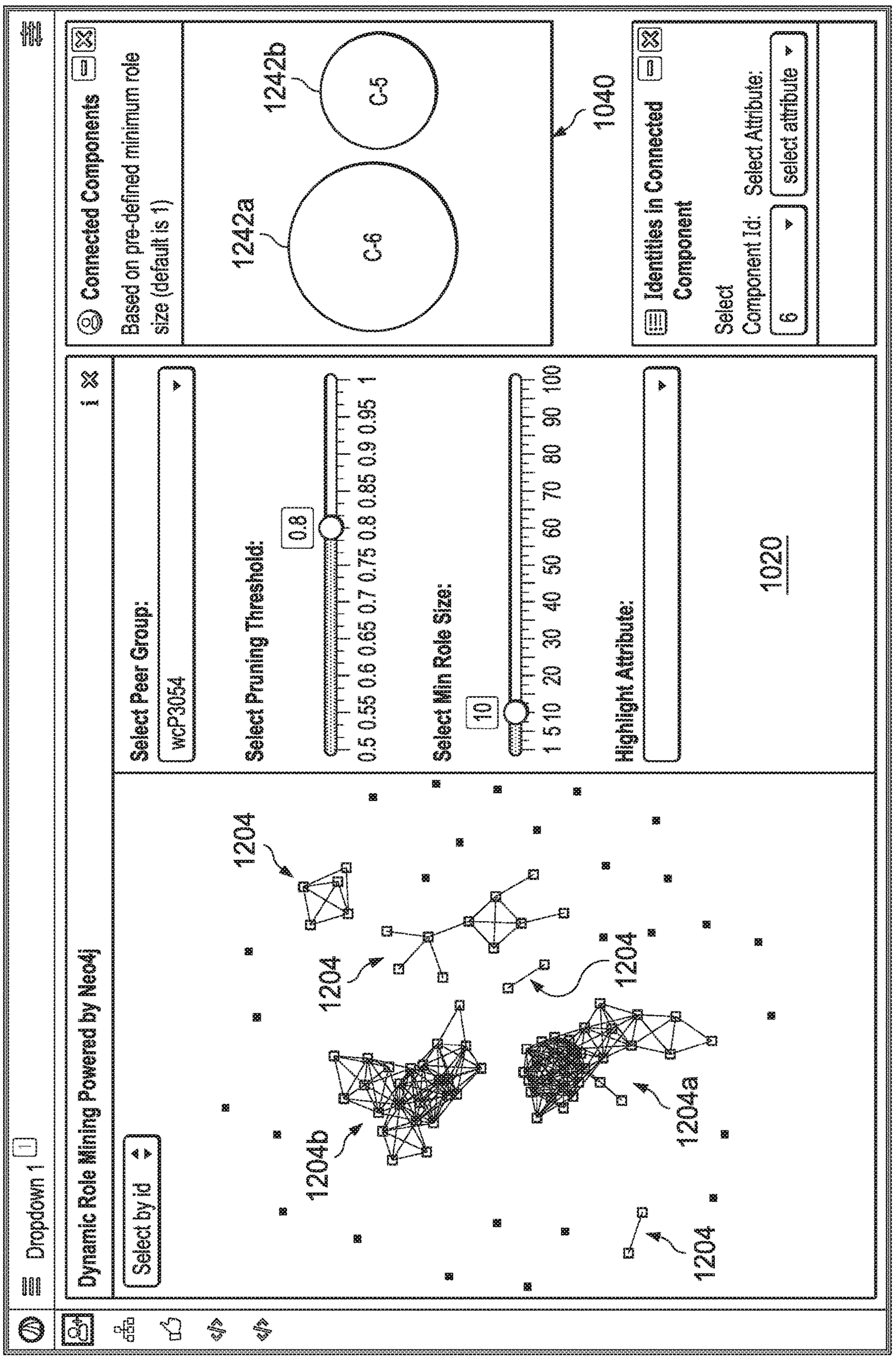
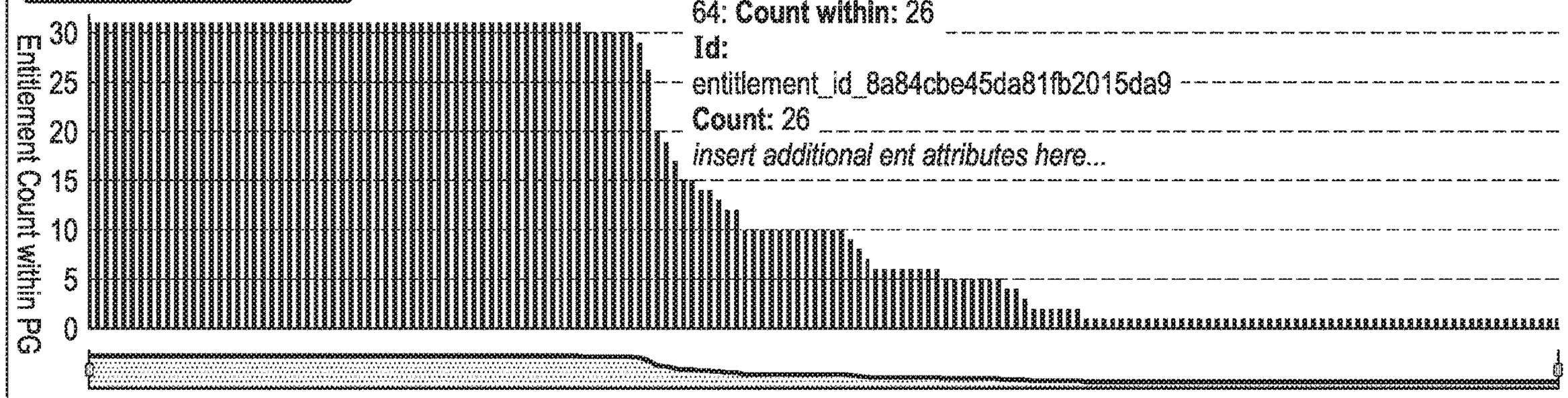

Connected Components
Based on pre-defined minimum role size (default is 1)
1242a
C-6
1242b
C-5
1040
Identities in Connected Component
Select Component Id:
6
Select Attribute:
select attribute
Select Peer Group:
wcP3054
Select Pruning Threshold:
0.8
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Select Min Role Size:
10
1 5 10 20 30 40 50 60 70 80 90 100
Highlight Attribute:
1020
Dropdown 1
Dynamic Role Mining Powered by Neo4j
Select by id
1204
1204
1204
1204
1204a
1204b

FIG. 12

Role-Entitlement Set Export

Role Evaluation

6



Show 10 entries

Search:

	e.id	count_inPG
1	entitlement_id_8a84cbe438adf1a801393db2417e2ded	31
2	entitlement_id_8a84cbe442c98dcc0142e45341901322	31
3	entitlement_id_8a84cbe442c98dcc0142df2c698c0bcd	31
4	entitlement_id_8a84cbe442c98dcc0142df2c66530bab	31
5	entitlement_id_8a84cbe459c344460159d42e4b7b05a2	31
6	entitlement_id_8a84cbe438adf1a801393df360c95a67	31
7	entitlement_id_8a84cbe442c98dcc0142e45345b01332	31
8	entitlement_id_8a84cbe45812772f01582662d86e67aa	31
9	entitlement_id_8a84cbe438adf1a801393df35f945a61	31
10	entitlement_id_8a84cbe43941c6970139c0e1f1367be5	31

Showing 1 to 10 of 169 entries

Previous 1 2 3 4 5 ... 17 Next

Please enter filename

Save to CSV

FIG. 13

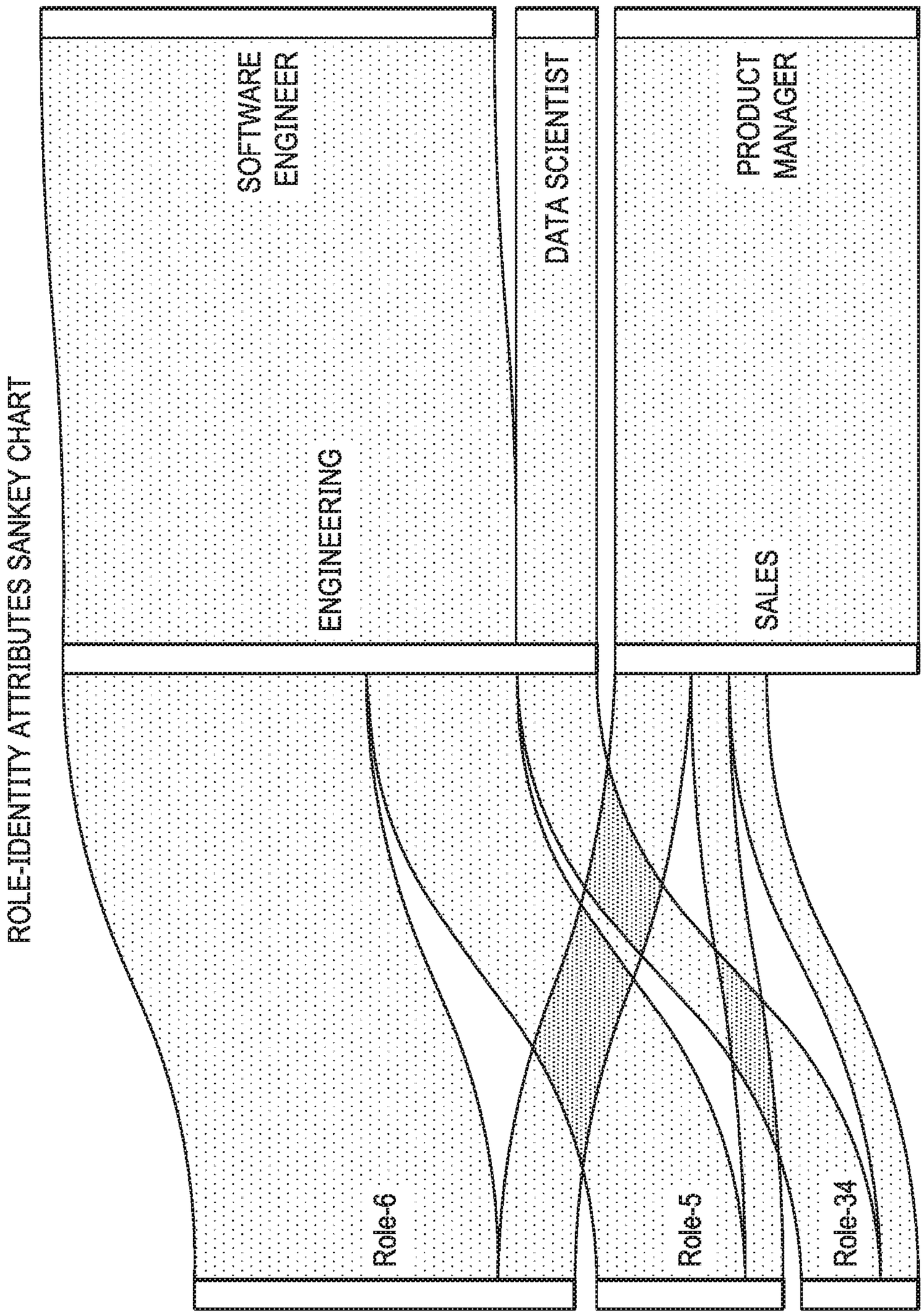

ROLE-IDENTITY ATTRIBUTES SANKEY CHART
SOFTWARE ENGINEER
DATA SCIENTIST
PRODUCT MANAGER
ENGINEERING
SALES
Role-6
Role-5
Role-34

FIG. 14

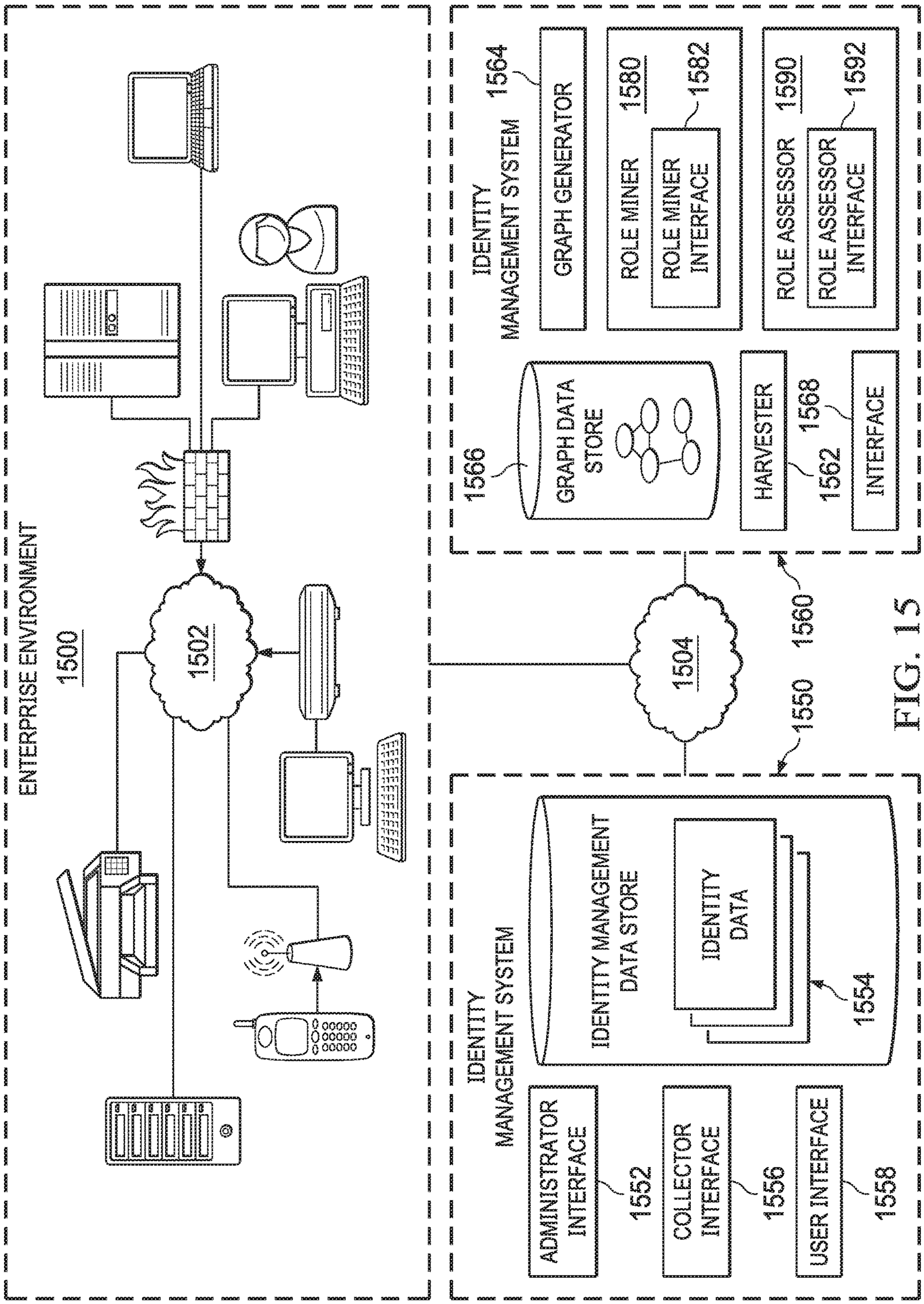

ENTERPRISE ENVIRONMENT
1500
1502
1504
1550
1560
IDENTITY MANAGEMENT SYSTEM
1564
GRAPH GENERATOR
ROLE MINER
1580
ROLE MINER INTERFACE
1582
ROLE ASSESSOR
1590
ROLE ASSESSOR INTERFACE
1592
1566
GRAPH DATA STORE
HARVESTER
1562
1568
INTERFACE
IDENTITY MANAGEMENT SYSTEM
IDENTITY MANAGEMENT DATA STORE
IDENTITY DATA
1554
ADMINISTRATOR INTERFACE
1552
COLLECTOR INTERFACE
1556
USER INTERFACE
1558

FIG. 15

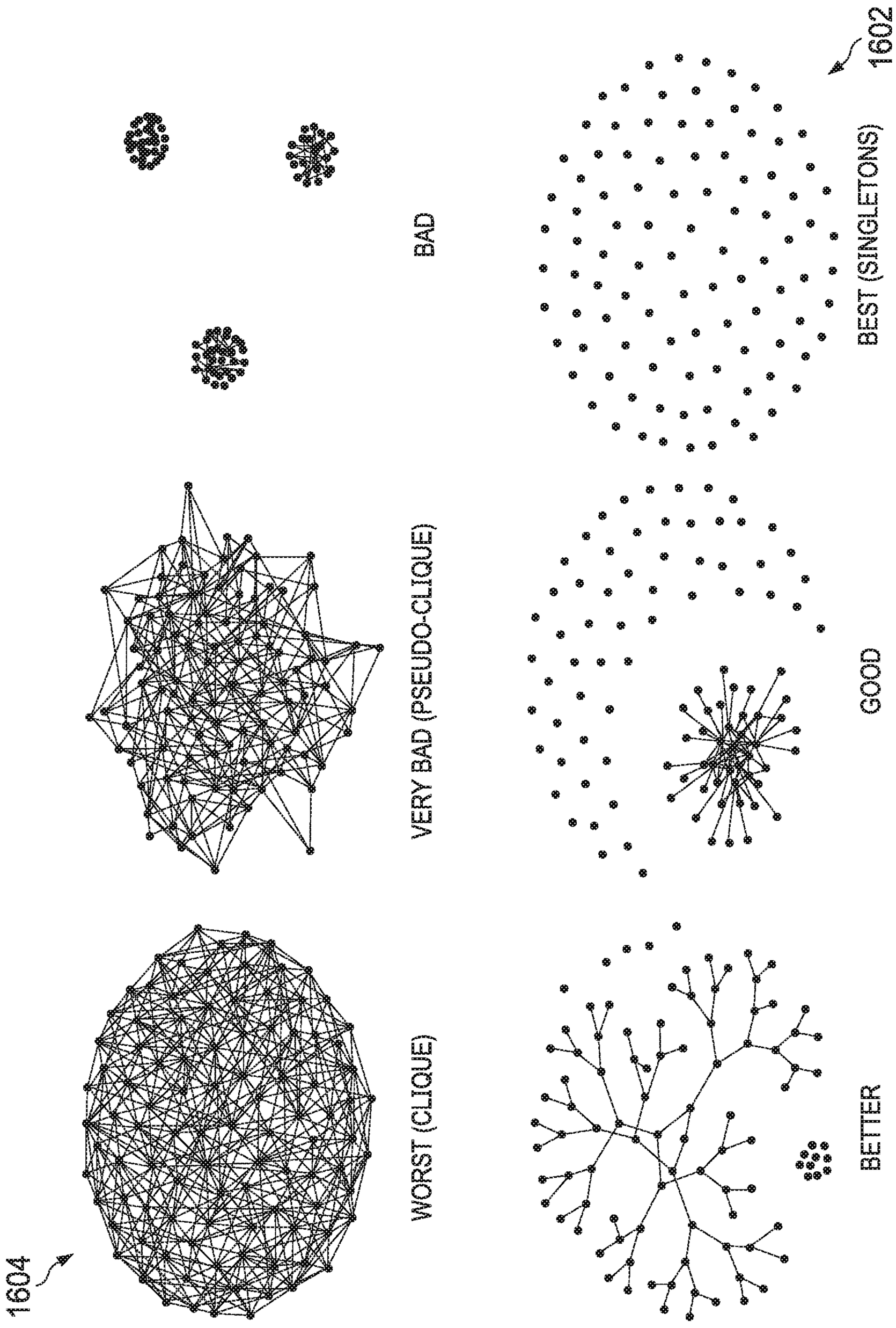

1602
BEST (SINGLETONS)
GOOD
BETTER
BAD
VERY BAD (PSEUDO-CLIQUE)
WORST (CLIQUE)
1604

FIG. 16

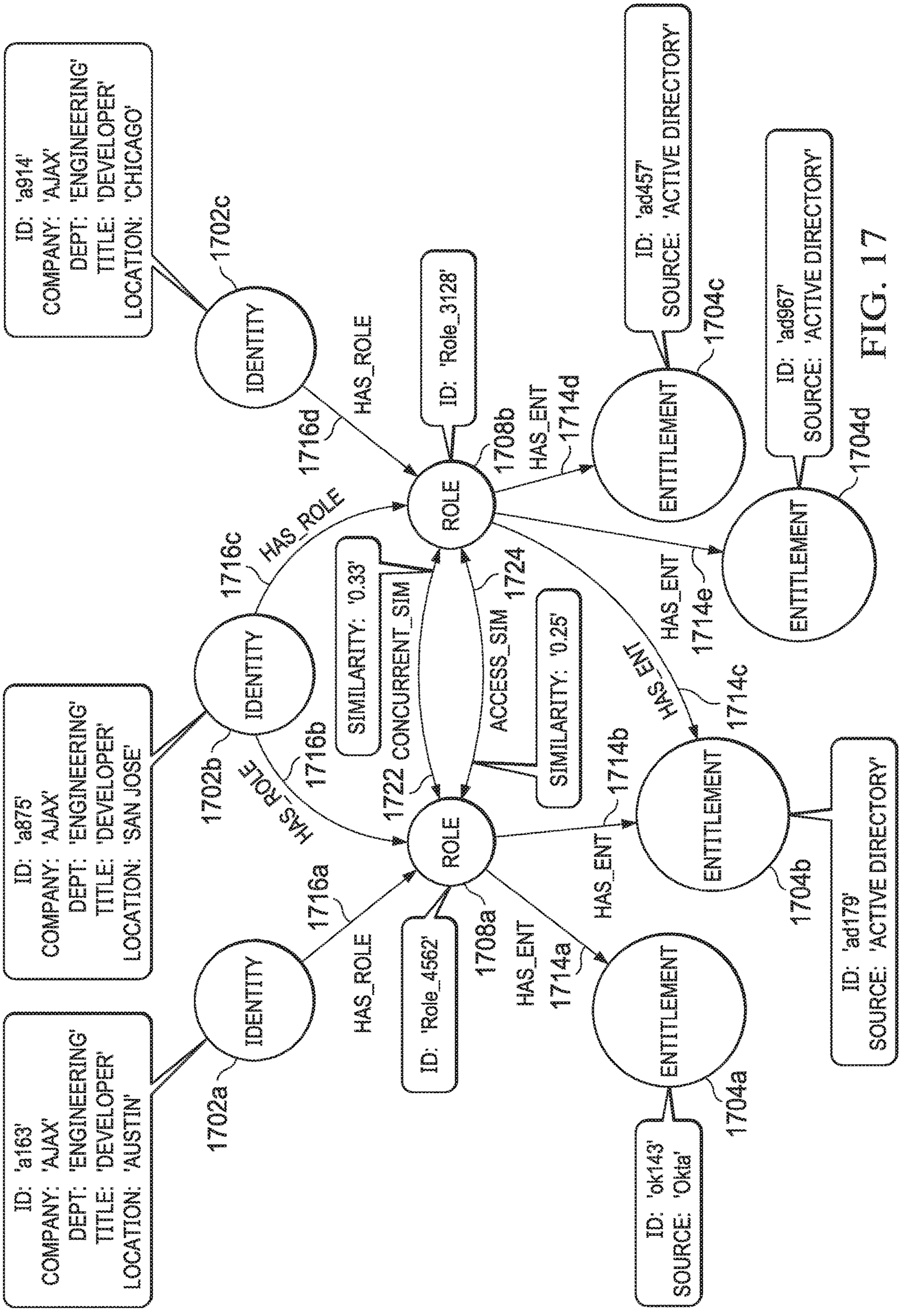

ID: 'a914'
COMPANY: 'AJAX'
DEPT: 'ENGINEERING'
TITLE: 'DEVELOPER'
LOCATION: 'CHICAGO'
1702c
IDENTITY
1716d
HAS_ROLE
ID: 'Role_3128'
ID: 'ad457'
SOURCE: 'ACTIVE DIRECTORY'
1704c
ID: 'ad967'
SOURCE: 'ACTIVE DIRECTORY'
ID: 'a875'
COMPANY: 'AJAX'
DEPT: 'ENGINEERING'
TITLE: 'DEVELOPER'
LOCATION: 'SAN JOSE'
1702b
IDENTITY
1716c
HAS_ROLE
ROLE
1708b
HAS_ENT
1714d
ENTITLEMENT
SIMILARITY: '0.33'
1722 CONCURRENT_SIM
ACCESS_SIM 1724
SIMILARITY: '0.25'
HAS_ENT
1714e
ENTITLEMENT
1716b
HAS_ROLE
ROLE
1714b
HAS_ENT
1714c
ENTITLEMENT
1716a
HAS_ROLE
ID: 'Role_4562'
1708a
HAS_ENT
1714a
ENTITLEMENT
ID: 'a163'
COMPANY: 'AJAX'
DEPT: 'ENGINEERING'
TITLE: 'DEVELOPER'
LOCATION: 'AUSTIN'
1702a
IDENTITY
ID: 'ok143'
SOURCE: 'Okta'
1704a
1704b
ID: 'ad179'
SOURCE: 'ACTIVE DIRECTORY'

FIG. 17

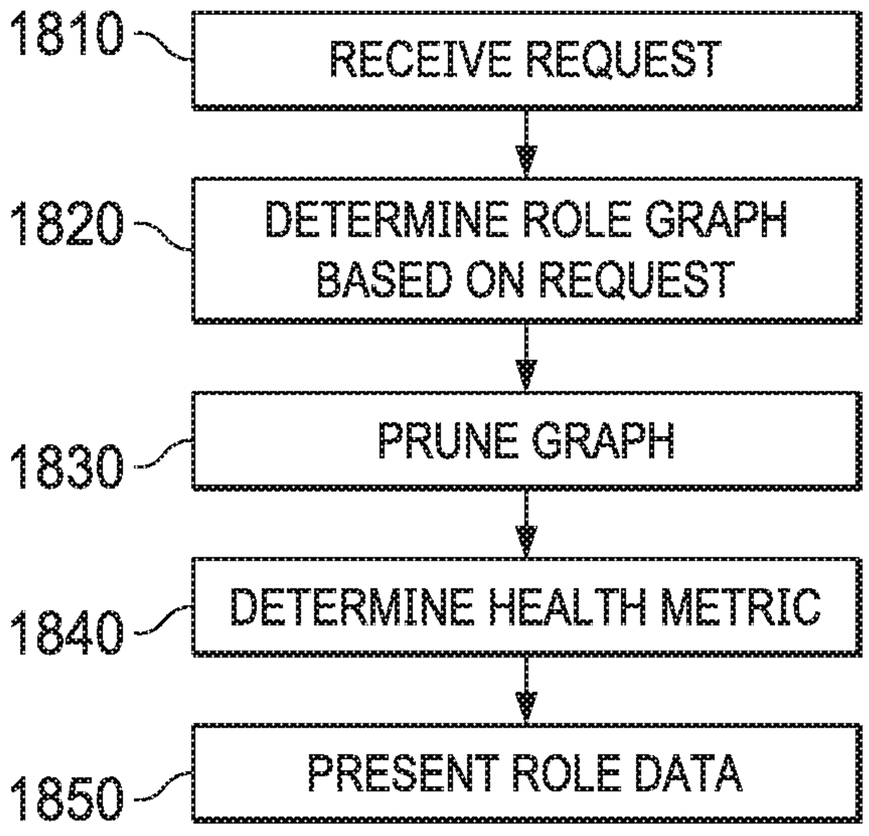

1810
RECEIVE REQUEST
1820
DETERMINE ROLE GRAPH BASED ON REQUEST
1830
PRUNE GRAPH
1840
DETERMINE HEALTH METRIC
1850
PRESENT ROLE DATA

FIG. 18

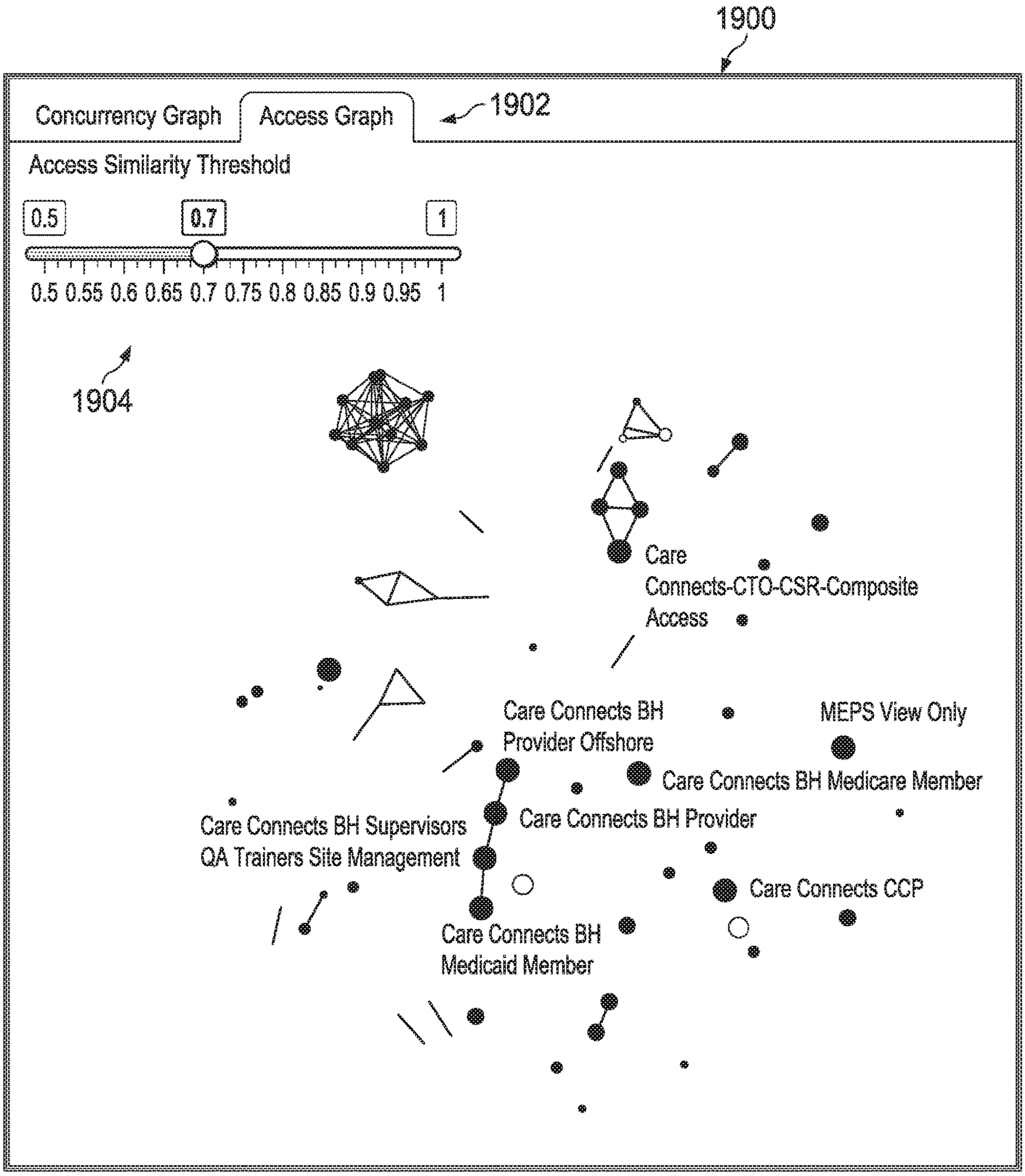

1900
Concurrency Graph
Access Graph
1902
Access Similarity Threshold
0.5
0.7
1
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
1904
Care Connects-CTO-CSR-Composite Access
Care Connects BH Provider Offshore
MEPS View Only
Care Connects BH Medicare Member
Care Connects BH Supervisors QA Trainers Site Management
Care Connects BH Provider
Care Connects CCP
Care Connects BH Medicaid Member

FIG. 19A

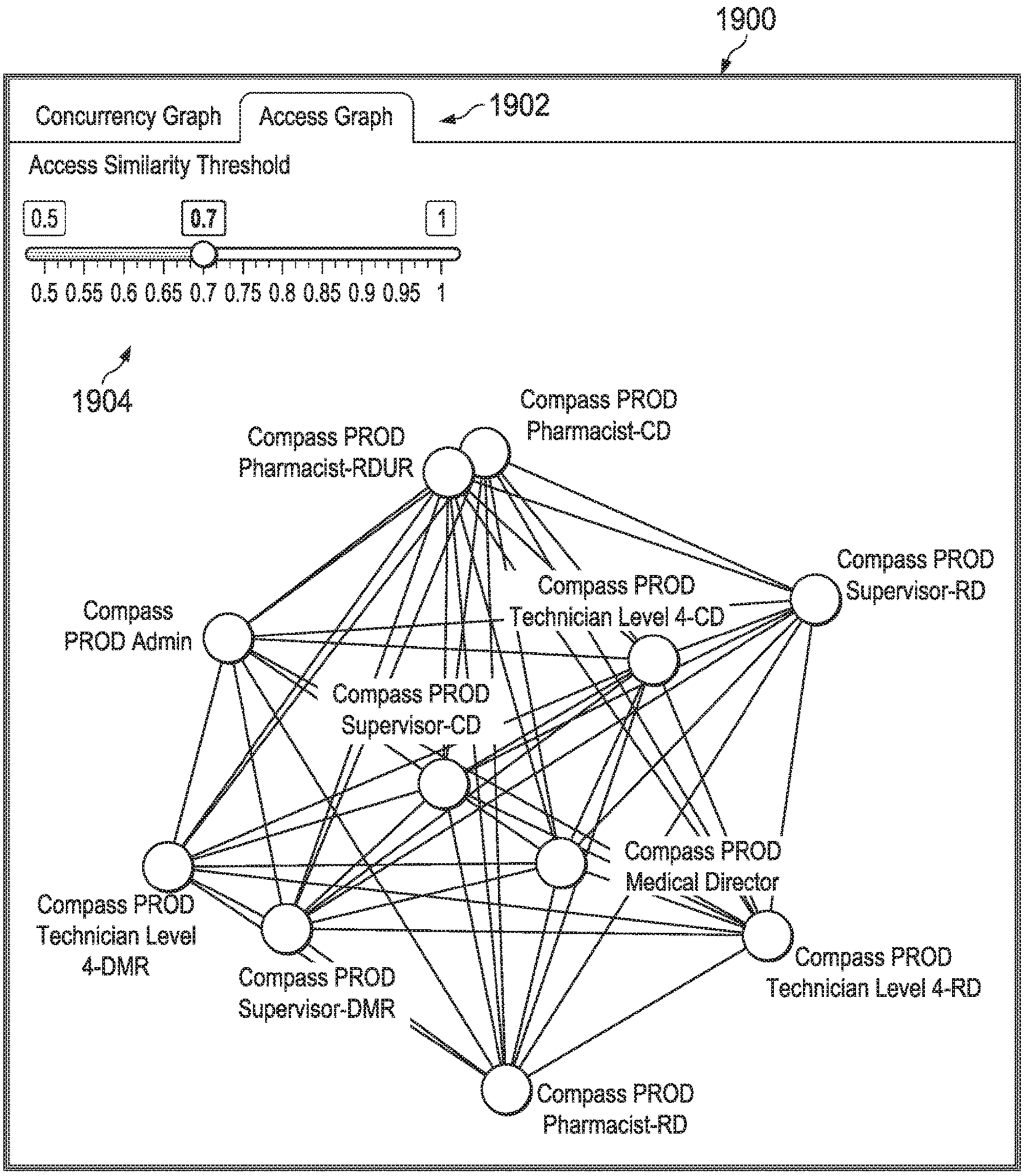

1900
Concurrency Graph
Access Graph
1902
Access Similarity Threshold
0.5
0.7
1
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
1904
Compass PROD Pharmacist-CD
Compass PROD Pharmacist-RDUR
Compass PROD Supervisor-RD
Compass PROD Technician Level 4-CD
Compass PROD Admin
Compass PROD Supervisor-CD
Compass PROD Medical Director
Compass PROD Technician Level 4-DMR
Compass PROD Technician Level 4-RD
Compass PROD Supervisor-DMR
Compass PROD Pharmacist-RD

FIG. 19B

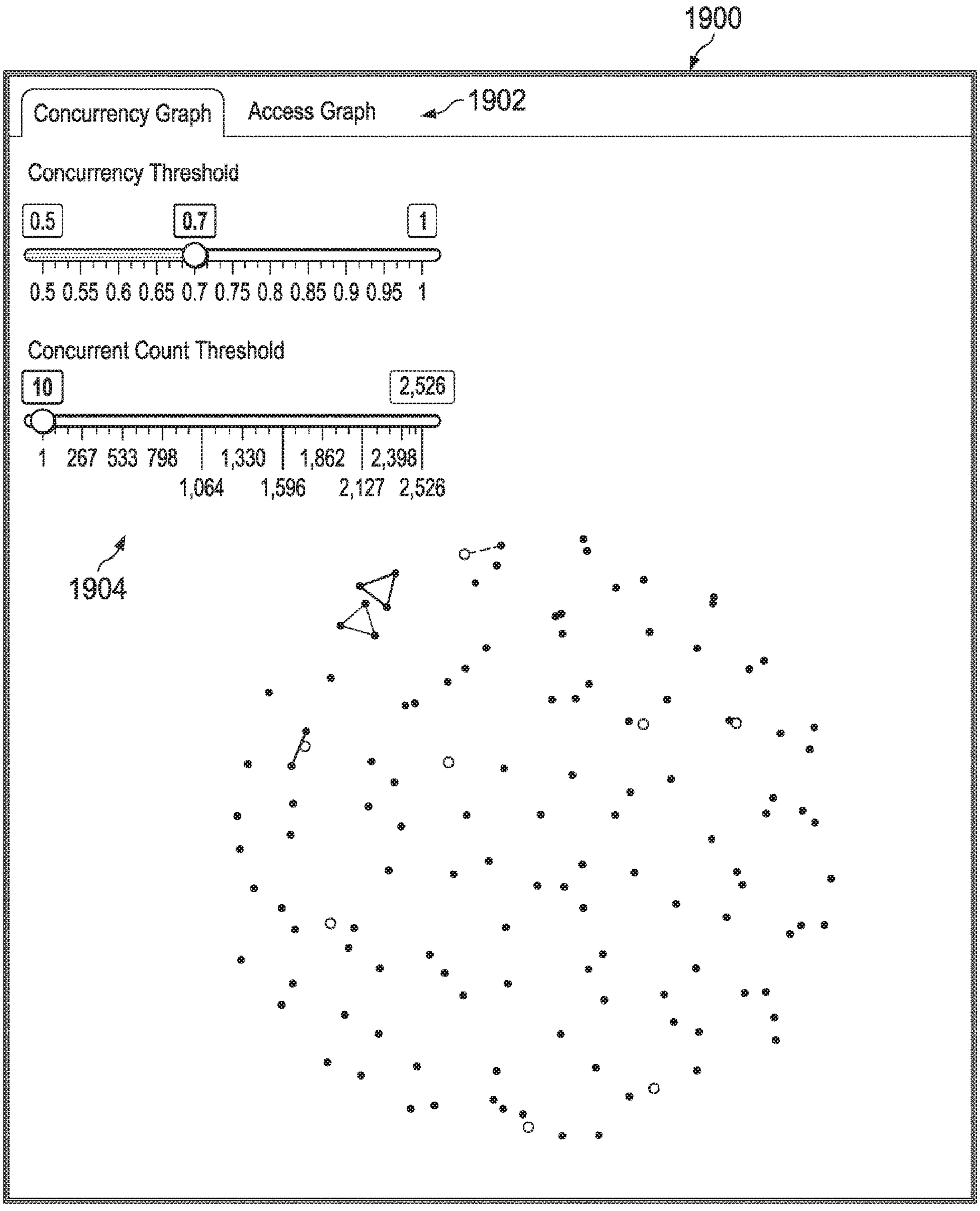

1900
Concurrency Graph
Access Graph
1902
Concurrency Threshold
0.5
0.7
1
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Concurrent Count Threshold
10
2,526
1 267 533 798 1,064 1,330 1,596 1,862 2,127 2,398 2,526
1904

FIG. 19C

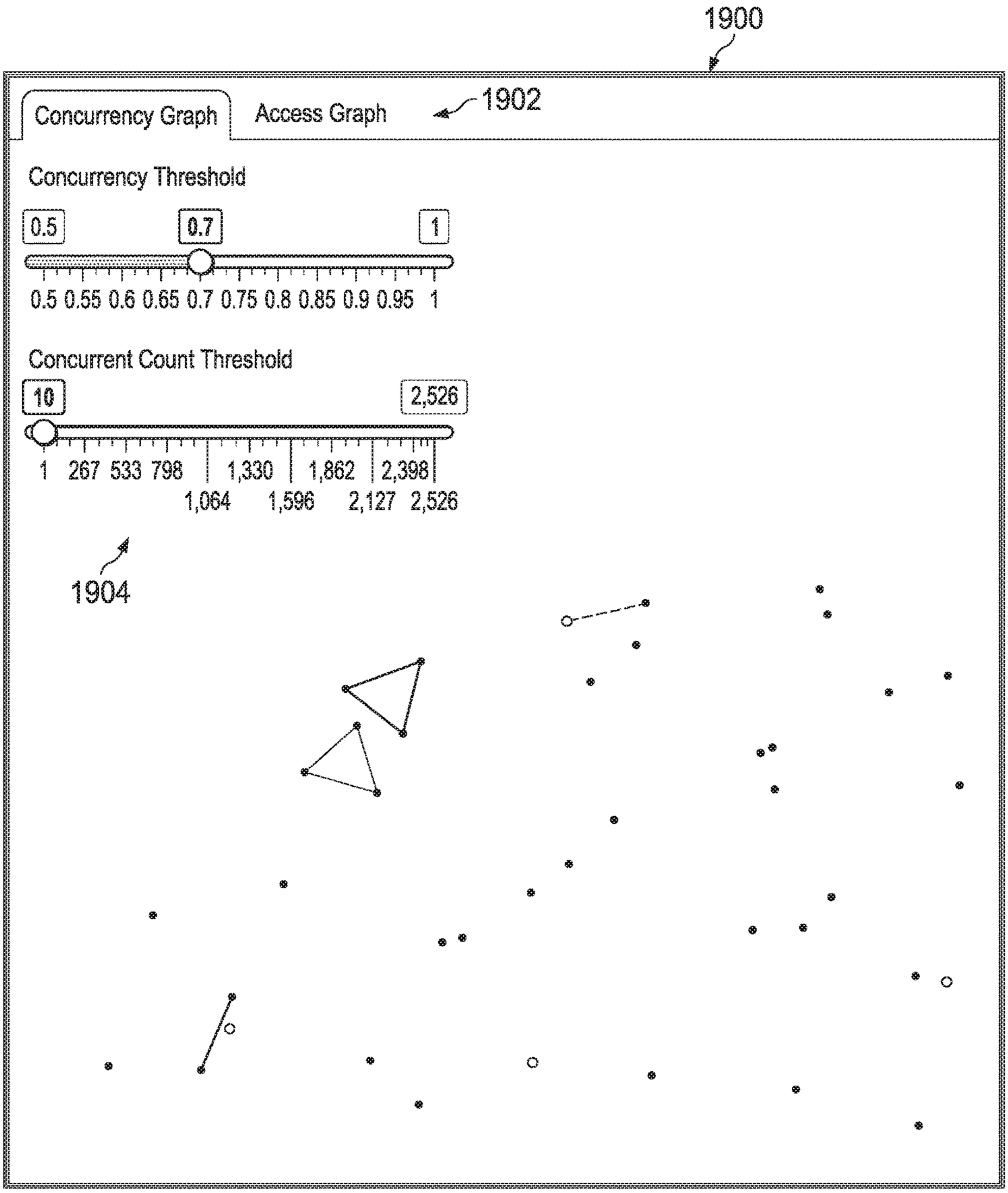

1900
Concurrency Graph
Access Graph
1902
Concurrency Threshold
0.5
0.7
1
0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Concurrent Count Threshold
10
2,526
1 267 533 798 1,064 1,330 1,596 1,862 2,127 2,398 2,526
1904

FIG. 19D

SYSTEM AND METHOD FOR ROLE VALIDATION IN IDENTITY MANAGEMENT ARTIFICIAL INTELLIGENCE SYSTEMS USING ANALYSIS OF NETWORK IDENTITY GRAPHS

COPYRIGHT NOTICE

TECHNICAL FIELD

This disclosure relates generally to computer security. In particular, this disclosure relates to the application of artificial intelligence to identity management in a distributed and networked computing environment. Even more specifically, this disclosure relates to enhancing computer security in a distributed networked computing environment through the use of role validation in these artificial intelligence based identity management systems, including the use of graph based analysis of roles and their associated entitlements or identities in association with such role validation.

BACKGROUND

Acts of fraud, data tampering, privacy breaches, theft of intellectual property, and exposure of trade secrets have become front page news in today's business world. The security access risk posed by insiders—persons who are granted access to information assets—is growing in magnitude, with the power to damage brand reputation, lower profits, and erode market capitalization.

Identity Management (IM), also known as Identity and Access Management (IAM) or Identity Governance (IG), is, the field of computer security concerned with the enablement and enforcement of policies and measures which allow and ensure that the right individuals access the right resources at the right times and for the right reasons. It addresses the need to ensure appropriate access to resources across increasingly heterogeneous technology environments and to meet increasingly rigorous compliance requirements. Escalating security and privacy concerns are driving governance, access risk management, and compliance to the forefront of identity management. To effectively meet the requirements and desires imposed upon enterprises for identity management, these enterprises may be required to prove that they have strong and consistent controls over who has access to critical applications and data. And, in response to regulatory requirements and the growing security access risk, most enterprises have implemented some form of user access or identity governance.

Yet many companies still struggle with how to focus compliance efforts to address actual risk in what usually is a complex, distributed networked computing environment. Decisions about which access entitlements are desirable to grant a particular user are typically based on the roles that the user plays within the organization. In large organizations, granting and maintaining user access entitlements is a difficult and complex process, involving decisions regarding whether to grant entitlements to thousands of users and hundreds of different applications and databases. This complexity can be exacerbated by high employee turnover, reorganizations, and reconfigurations of the various accessible systems and resources.

Organizations that are unable to focus their identity compliance efforts on areas of greatest access risk can waste time, labor, and other resources applying compliance monitoring and controls across the board to all users and all applications. Furthermore, with no means to establish a baseline measurement of identity compliance, organizations have no way to quantify improvements over time and demonstrate that their identity controls are working and effectively reducing access risk.

Information Technology (IT) personnel of large organizations often feel that their greatest security risks stemmed from "insider threats," as opposed to external attacks. The access risks posed by insiders range from careless negligence to more serious cases of financial fraud, corporate espionage, or malicious sabotage of systems and data. Organizations that fail to proactively manage user access can face regulatory fines, litigation penalties, public relations fees, loss of customer trust, and ultimately lost revenue and lower stock valuation. To minimize the security risk posed by insiders (and outsiders), business entities and institutions alike often establish access or other governance policies that eliminate or at least reduce such access risks and implement proactive oversight and management of user access entitlements to ensure compliance with defined policies and other good practices.

One of the main goals of identity management, then, is to help users identify and mitigate risks associated with access management. Many times this access risk may result as an outgrowth of the evolution of roles within an enterprise over time. As roles have entitlements added or deleted and as different roles are assigned or removed from different identities these changes may create a complex system that evolves in unpredictable ways over time. As the roles and identities evolve, they may stray in substantial and detrimental ways from the 'gold standard' of the role definition or other identity governance desires of the enterprise. While some enterprises manage to iteratively engineer or re-design or re-define their role structures and access model to keep pace with security requirements, the majority of enterprises are unaware of the efficacy of their access security due to the lack of abilities to monitor and evaluate efficacy of overall access landscape, especially in the context of roles defined for, or utilized by, the enterprise.

Accordingly, it is desirable for identity management solutions to offer tools to assist in the assessment of roles or other identity management artifacts associated with the identity management data associated with enterprise.

SUMMARY

As mentioned, it is desirable for identity management solutions to offer role assessment capabilities whereby roles comprising collections of entitlements may be ascertained from the identity management data associated with enterprise and an assessment metric (also refer to as a score) for a set of these roles may be determined, where the metric is a reflection, for example, of the quality or health of the structure of the set of roles. Specifically, in many instances, in the context of an enterprise there may be what are referred to as multi-dimensional roles. A multi-dimensional role may be instances of similar roles that may vary slightly according to some criteria. For example, if an enterprise has many different locations, a role in one location (e.g., a software developer role in Austin, Tex.) may be very similar to a role

in another location (a role for a software developer in San Jose, Calif.). In other words, a software developer in either location may require access to a substantially similar set of entitlements, however, since the creators of such roles (which may be, for example, in those two different locations) may have no visibility or access into the roles structure of the enterprise generally, two (or more) different roles may be created, despite the fact that these roles may be substantially similar (e.g., comprise similar entitlements) or, in certain cases, may even be the same. Thus, administrators or others concerned with identify governance within an enterprise, or compliance of an enterprise with identity management goals or requirements, may desire to validate or otherwise assess the role structure of an enterprise (or portions thereof) to determine the quality or health of these roles. By assessing the health of the roles structure, such metrics may be useful for compliance purposes or to assist in optimizing the role structure or more generally streamlining role management for the enterprise.

What is desired, therefore, are effective system and methods for providing a holistic view and assessment of the overall access model health across an enterprise, and specifically for assessing the health of role structures within an enterprise.

By identifying roles that may be strongly similar or otherwise closely aligned, efficiencies with respect to management of these roles may be achieved. For example, in some cases, roles that have similar sets of entitlements may be consolidated (e.g., merged) or some of the roles eliminated. As another alternative, roles that share a similar group of identities (e.g., where the same set of identities share a set of roles) may be bundled together and an overarching role (referred to as a portfolio role) may be defined such that the bundle of similar roles may be manage as a group using the portfolio role. Thus, using embodiments, the actual scope of identities (e.g., a user population) for which roles can be consolidated to reduce use of resources in role management for that specific population and defining or assigning roles for that population. More generally, then, by reducing the number of roles or the interactions with these roles, the number of both computing resources and man hours required for such identity governance may be reduced, along with the commensurate cost to the enterprise of such identity management.

To those ends, among others, attention is now directed to the embodiments of artificial intelligence based identity governance systems that provide such role assessment. Accordingly, to ameliorate or address these issues, among other ends, embodiments of the identity management systems disclosed herein may utilize a network graph approach to improve identity governance, including the assessment of roles associated with the identity management data of an enterprise. Specifically, embodiments of identity management systems as disclosed may provide role assessment based on a network graph that includes roles of an enterprise. Embodiments may thus generate a network identity graph that includes nodes for identities, entitlements, roles or other identity management artifacts of an enterprise. Such a network identity graph may be, or may include, a role graph having nodes representing roles associated with the enterprise and edges representing similarities between the roles (e.g., represented by the nodes). These edges may comprise a similarity weight determined, based on, for example, shared entitlements between the roles or by concurrent identities (e.g., a number of identities that share those roles).

In one embodiment, for example, the role graph may be an access role graph that is a role graph modeled in terms of entitlement (e.g., access) similarities between all the roles. The edges of the access role graph represent an access similarity relationship between two roles (e.g., nodes representing the roles) joined by the edge of the graph. A weight may be computed for the access similarity relationship based on the entitlements shared between the two roles and the number of entitlements the roles include. Roles with similar entitlements or access patterns may thus cluster close together on the access graph. Embodiments of these access role graphs may give high-level of abstractions on the overall access model of an enterprise while accurately reflecting the global role (access) structure. As such, these access role graphs may be useful, for example, as a role provisioning QA (Quality Assessment) tool indicating overall well-being of an enterprises role structure, in recommending consolidation of redundant roles, or verifying how new roles may fit in the current access model.

As another embodiment, for example, the role graph may be a concurrency role graph (also referred to as a concurrency or concurrent graph) that is modeled in terms of concurrent identities shared between roles. The edges of the concurrency graph represent an concurrency similarity relationship between two roles (e.g., nodes representing the roles) joined by the edge of the graph. A weight may be computed for the concurrency similarity relationship based on the number of identities which share those roles and the number of identities that have those roles. Roles with high concurrency with one another cluster closer together on the concurrency graph. Moreover, the concurrency graph may be filtered based on the number of supporting identities (e.g., the number of identities that include both roles). This support thus determines the significance of the computed concurrency weights, by allowing the concurrency graph to filtered to filter out highly concurrent roles that share only few identities, thus rendering more meaningful representation of the concurrency graph. As such, these concurrency graphs may be useful as a "role-profiling assistant" identifying concurrent patterns of peer access, simplifying business rules, or surfacing potential profiles for new joiners. These concurrency graphs may also allow users to dive deeper and profile roles within units of an enterprise when applied with scoping of the concurrency graph.

Moreover, according to embodiments, various metrics may be determined for assessing the quality or health of the role structure of an enterprise based on an access role graph or a concurrency role graph. Specifically, optimal (e.g., ideal) network or graph topologies for access and concurrency graphs can be inferred. Graph based metrics may thus provide a starting point to standardize quality scoring for role structures and access models. In one embodiment, a combination of graph based metrics may be utilized to measure a role graph structure with respect to an ideal graph topology optimized for the enterprise. Such a scoring system allows personalization taking into account the trade-off between compliance-driven and enablement-driven governance strategies. Thus role data, including for example, visual depictions of role graphs for the enterprise or quality assessment scores may be presented to a user through embodiments of the identity management systems as depicted herein.

Enterprises (e.g., with existing access governance strategy) could thus utilize a role validation or access awareness interface presented by the identity management system to evaluate and validate their existing role structure to, for example, explore hierarchical relationships between existing

roles, profile, re-provision, or label (e.g., tag) highly similar existing roles (e.g., similarity 75% or more), consolidate and label existing roles that are heavily concurrent within certain populations, or evaluate the health of an entire (or portion of a) role structure based on the scoring system or visual depiction of a role graph.

Similarly, enterprise involved in in active access modeling or governance process (e.g., using role mining capabilities) could utilize the role validation or access awareness interface for decisions related to prioritizing roles based on the novelty with respect to existing roles; the provisioning of newly discovered roles with significantly high contrast to existing roles; validating the impact of provisioned roles under a current role structure, merging, profiling, or labeling (tag) highly similar existing roles, or enhancing access interpretability and enabling detection of potential risk based on security policies.

Embodiments provide numerous advantages over previously available systems and methods for measuring access risk. As embodiments are based on a graph representation of identity management data, the graph structure may serve as a physical model of the data, allowing more intuitive access to the data (e.g., via graph database querying, or via graph visualization techniques). This ability may yield deeper and more relevant insights for users of identity management systems. Such abilities are also an outgrowth of the accuracy of the results produced by embodiments as disclosed.

As such, these embodiments of identity management systems may allow an accurate approach to role validation in identity governance. This will allow the identification and assessment of the role structure of an enterprise roles and the evolution of such a role structure. Ultimately, this will yield an improved model for roles that will accurately match the evolving access entitlement system.

Moreover, the graph format used by certain embodiments, allows the translation of domain and enterprise specific concepts, phenomena, and issues into tangible, quantifiable, and verifiable hypotheses which may be examined or validated with graph-based algorithms. Accordingly, embodiments may be especially useful in assessing risk and in compliance with security policies or the like.

Additionally, embodiments as disclosed may offer the technological improvement of reducing the computational burden and memory requirements of systems implementing these embodiments through the improved data structures and the graph processing and analysis implemented by such embodiments. Accordingly, embodiments may improve the performance and responsiveness of identity management systems that utilize such embodiments of identity graphs and clustering approaches by reducing the computation time and processor cycles required (e.g., and thus improving processing speed) and simultaneously reducing memory usage or other memory requirements.

These, and other, aspects of the disclosure will be better appreciated and understood when considered in conjunction with the following description and the accompanying drawings. It should be understood, however, that the following description, while indicating various embodiments of the disclosure and numerous specific details thereof, is given by way of illustration and not of limitation. Many substitutions, modifications, additions and/or rearrangements may be made within the scope of the disclosure without departing from the spirit thereof, and the disclosure includes all such substitutions, modifications, additions and/or rearrangements.

BRIEF DESCRIPTION OF THE FIGURES

The drawings accompanying and forming part of this specification are included to depict certain aspects of the invention. A clearer impression of the invention, and of the components and operation of systems provided with the invention, will become more readily apparent by referring to the exemplary, and therefore nonlimiting, embodiments illustrated in the drawings, wherein identical reference numerals designate the same components. Note that the features illustrated in the drawings are not necessarily drawn to scale.

FIG. **1** is a block diagram of a distributed networked computer environment including one embodiment of an identity management system.

FIG. **2** is a flow diagram of one embodiment of a method for peer group detection and analysis using cluster based analysis of identity graphs.

FIGS. **3**A, **3**B, **3**C, **3**D and **3**E depict example visual representations of identity graphs.

FIGS. **4-7** depict interfaces that may be utilized by embodiments of an identity management system.

FIG. **8** is a block diagram of a distributed networked computer environment including one embodiment of an identity management system.

FIG. **9** is a flow diagram of one embodiment of a method for role mining.

FIGS. **10-14** depict interfaces that may be utilized by embodiments of an identity management system.

FIG. **15** is a block diagram of a distributed networked computer environment including one embodiment of an identity management system.

FIG. **16** depict example visual representations of role graphs.

FIG. **17** depicts an example representation of a role graph.

FIG. **18** is a flow diagram of one embodiment of a method for role assessment.

FIGS. **19**A, **19**B, **19**C and **19**D depict interfaces that may be utilized by embodiments of an identity management system.

DETAILED DESCRIPTION

The invention and the various features and advantageous details thereof are explained more fully with reference to the non-limiting embodiments that are illustrated in the accompanying drawings and detailed in the following description. Descriptions of well-known starting materials, processing techniques, components and equipment are omitted so as not to unnecessarily obscure the invention in detail. It should be understood, however, that the detailed description and the specific examples, while indicating some embodiments of the invention, are given by way of illustration only and not by way of limitation. Various substitutions, modifications, additions and/or rearrangements within the spirit and/or scope of the underlying inventive concept will become apparent to those skilled in the art from this disclosure.

Before delving into more details regarding the specific embodiments disclosed herein, some context may be helpful. In response to regulatory requirements and security access risks and concerns, most enterprises have implemented some form of computer security or access controls. To assist in implementing security measures and access controls in an enterprise environment, many of these enterprises have implemented Identity Management in association with their distributed networked computer environments. Identity Management solutions allow the definition of a function or an entity associated with an enterprise. An identity may thus represent almost physical or virtual entity, place, person or other item that an enterprise would like to define. Identities can therefore represent, for example, func-

tions or capacities (e.g., manager, engineer, team leader, etc.), title (e.g., Chief Technology Officer), groups (development, testing, accounting, etc.), processes (e.g., nightly back-up process), physical locations (e.g., cafeteria, conference room), individual users or humans (e.g., John Locke) or almost any other physical or virtual entity, place, person or other item. Each of these identities may therefore be assigned zero or more entitlements with respect to the distributed networked computer environments. An entitlement may be the ability to perform or access a function within the distributed networked computer environments, including, for example, accessing computing systems, applications, file systems, particular data or data items, networks, subnetworks or network locations, etc.

To facilitate the assignment of these entitlements, enterprises may also be provided with the ability to define roles within the context of their Identity Management solution. A role within the context of Identity Management may be a collection of entitlements. These roles may be assigned a name or identifiers (e.g., manager, engineer, team leader) by an enterprise that designate the type of user or identity that should be assigned such a role. By assigning a role to an identity in the Identity Management context, the identity may be assigned the corresponding collection of entitlements associated with the assigned role. Accordingly, by defining these roles enterprises may define a "gold standard" of what they desire their identity governance to look like.

Thus, by managing the roles within the enterprise computing environment, the assignment of entitlements and the proliferation of these roles or entitlements may be controlled. However, escalating security and privacy concerns are driving governance, access risk management, and compliance to the forefront of Identity Management. Yet many companies still struggle with how to focus compliance efforts to address actual risk in what usually is a complex, distributed networked computing environment. Decisions about which access roles or entitlements are desirable to grant a particular user are typically based on the business roles that the user plays within the organization. In large organizations, granting and maintaining roles and user access entitlements is a difficult and complex process, involving decisions regarding whether to grant roles or entitlements to thousands of users and hundreds of different applications and databases. This complexity can be exacerbated by high employee turnover, reorganizations, and reconfigurations of the various accessible systems and resources.

However, to effectively meet the requirements and desires imposed upon enterprises for Identity Management, these enterprises may be required to prove that they have strong and consistent controls over who has access to critical applications and data. Generally then, what is desired are effective system and methods for providing a holistic view and assessment of the overall access model health across an enterprise, and specifically for assessing the health of role structures within an enterprise. More specifically, it is desirable for identity management solutions to offer role assessment capability whereby roles may be ascertained from the identity management data associated with enterprise and the structure of these roles assessed or presented to a user. Additionally, it may be desirable to present an assessment metric for these roles, where the metric is a reflection, for example, of the quality or health of the structure of the set of roles.

To those ends, among others, attention is now directed to the embodiments of artificial intelligence based identity governance systems that provide such role assessment. Specifically, embodiments of the identity management systems disclosed herein may utilize a network graph approach to improve identity governance, including the assessment of roles associated with the identity management data of an enterprise. In particular, embodiments of identity management systems as disclosed may provide role assessment based on a network graph that includes roles of an enterprise. Embodiments may thus generate a network identity graph that includes nodes for identities, entitlements, roles or other identity management artifacts of an enterprise. Such a network identity graph may be, or may include, a role graph having nodes representing roles associated with the enterprise and edges representing similarities between the roles (e.g., represented by the nodes). These edges may comprise a similarity weight determined, based on, for example, shared entitlements between the roles or by concurrent identities (e.g., a number of identities that share those roles).

In one embodiment, for example, the role graph may be an access role graph that is a role graph modeled in terms of entitlement (e.g., access) similarities between all the roles. A weight may be computed for the access similarity relationship based on the entitlements shared between the two roles and the number of entitlements the roles include. Embodiments of these access role graphs may give high-level of abstractions on the overall access model of an enterprise while accurately reflecting the global role (access) structure. As such, these access role graphs may be useful, for example, as a role provisioning QA (Quality Assessment) tool indicating overall well-being of an enterprise's role structure, in recommending consolidation of redundant roles, or verifying how new roles may fit in the current access model.

As another embodiment, for example, the role graph may be a concurrency role graph (also referred to as a concurrency or concurrent graph) that is modeled in terms of concurrent identities shared between roles. A weight may be computed for the concurrency similarity relationship based on the number of identities which share those roles and the number of identities that have those roles. These concurrency graphs may be useful as a "role-profiling assistant" identifying concurrent patterns of peer access, simplifying business rules, or surfacing potential profiles for new joiners. These concurrency graphs may also allow users to dive deeper and profile roles within units of an enterprise when applied with scoping of the concurrency graph.

Moreover, according to embodiments, various metrics may be determined for assessing the quality or health of the role structure of an enterprise based on an access role graph or a concurrency role graph. Specifically, optimal (e.g., ideal) network or graph topologies for access and concurrency graphs can be inferred. Graph based metrics may thus provide a starting point to standardize quality scoring for role structures and access models. In one embodiment, a combination of graph based metrics may be utilized to measure a role graph structure with respect to an ideal graph topology optimized for the enterprise. Such a scoring system allows personalization taking into account the trade-off between compliance-driven and enablement-driven governance strategies. Thus role data, including for example, visual depictions of role graphs for the enterprise or quality assessment scores may be presented to a user through embodiments of the identity management systems as depicted herein.

Embodiments as disclosed herein may thus provide role assessment from an enterprise's actual identity management data. By determining a current snapshot of the roles mined

from an actual state of the enterprise's identity governance structure, the enterprise roles as defined by the users of the enterprise may be compared with a desired state of the roles to reduce discrepancies therebetween, including for example, the identification of new roles, the evolution of the enterprise defined roles to match the evaluation of the actual role structure (e.g., the assessed roles), or the performance housekeeping on the assignment of entitlements or roles within the enterprise to more particularly tailor the actual role structure to an ideal role structure.

In certain cases, the efficacy of embodiments of role assessment in an identity management system may depend at least partially on the state of the identities, entitlements or roles within a distributed computing enterprise. Accordingly, before embodiments of the role assessment are discussed in more detail, it may be useful to an understanding of certain embodiments if the analysis and use of roles, entitlements and identities of an enterprise by embodiments of artificial intelligence identity governance systems are discussed in more detail, as such data may be used in the role assessment itself.

With that in mind, it may be understood that good governance practice in the identity space relies on the 'social' principle that identities with strongly similar attributes should be assigned similar, if not identical, access entitlements. In the realm of identity governance and administration, this approach allows for a separation of duties and thus makes it feasible to identify, evaluate, and prioritize risks associated with privileged access. As part of a robust identity management system, it is therefore highly desirable to analyze an enterprise's data to identify potential risks. In principle, strictly enforced pre-existing governance policies should ensure that identities with strongly similar access privileges are strongly similar. It would thus be desirable to group or cluster the identities of an enterprise into peer groups such that the identities in a peer group are similar with respect to the set of entitlements assigned to the identities of that group (e.g., relative to other identities or other groups). Peer grouping of the identities within an enterprise (or viewing the peer groups of identities) may allow, for example, an auditor or other person performing a compliance analysis or evaluation to quantitatively and qualitatively assess the effectiveness of any applicable pre-existing polices, or lack thereof, and how strictly they are enforced.

However, the data utilized by most identity management systems is not strictly numerical data. Often this data includes identifications of identities (e.g., alphanumeric identifiers for an identity as maintained by an identity management system) and identifications of entitlements or roles associated with those identities (e.g., alphanumeric identifiers for entitlements or roles as maintained by the identity management system). This data may also include data identifying roles (e.g., alphanumeric identifiers or labels for a role as maintained by an identity management solution) and identifications of entitlements associated with those roles (e.g., alphanumeric identifiers for the collection of entitlements associated with those roles). Clustering of this type of categorical data (e.g., for peer grouping of identities) is typically a harder task than clustering data of numerical type. In particular, clustering categorical data is particularly challenging since intuitive, geometric-based, distance measures experienced in real life, e.g., Euclidean distance, by definition, are exclusive to numerical data. A distance measure is a crucial component of any clustering algorithm as it is utilized at the lowest level to determine how similar/dissimilar two data points are.

For example, the one-hot-encoding data transform, which can convert categorical data into numerical data, does not work in these types of cases. Due to large number of entitlements, when combining the numerical, high-dimensional, one-hot encoded data with traditional geometric distances (e.g., Euclidean), distances between data points will be quite large and will make it hard, if not impossible, for a clustering algorithm to yield meaningful outputs. This is a direct mathematical outcome to the high dimensionality of the ambient space. It is a well-documented issue in data science literature, and the applicable nomenclature is "curse of dimensionality". Typical dimensionality reduction techniques (e.g., PCA, t-SNE, etc.) have been experimented with, but due to the way these clustering algorithms manipulate numerical data, the resulting transforms may manipulate the original data in ways that are not interpretable, hence not useful in this context.

Accordingly, conventional statistical clustering such as K-modes, or K-modes used in association with a data-mining, pattern-finding algorithm such as Equivalence Class Transformation (ECLAT), have thus proven inadequate. Many of the reasons for the inadequacy of such typical clustering approaches have to do with the computationally intensive nature of the computer implementations of such clustering, which are both computationally and memory intensive, reducing or hindering the performance and responsiveness of identity management systems that utilize such clustering approaches.

Attempts to remedy these problems by altering the clustering to discard or ignore less popular identities or entitlements to enhance the signal-to-noise ratio in their application have been less than successful, achieving neither adequate results in the clusters determined or in improving the performance or memory usage of systems which employ such clustering. Other workarounds for these deficiencies have also proven unworkable to this type of identity and entitlement data.

Moreover, when attempting to cluster based on categorical data, typical clustering algorithms do not capture the social aspects of identity governance. Homophily in social networks, as defined in social sciences, is the tendency of individuals to associate and bond with similar others. In identity governance, homophily in the identity space usually results as a consequence of enforcing the governance principle that similar identities should be assigned similar access entitlements. It is thus important to attempt to capture, or otherwise utilize this homophily, when peer grouping for identity management. As a consequence of all these deficiencies, the results from prior approach to identity clustering in the context of identity management were harder to interpret, yielding fewer insights, and negatively impacting the performance, efficiency, and overall quality of identity management systems. The data-driven clustering approach of identities into peer groups remains, however, a crucial component of identity management in a distributed and networked computing environment for a variety of reasons, including the usefulness of reviewing and visualizing such clusters of identities for auditing and compliance purposes.

Accordingly, to ameliorate these issues, among other ends, embodiments of the identity management systems disclosed herein may utilize a network graph approach to peer grouping of identities and entitlements of distributed networked enterprise computing environment. Specifically, in certain embodiments, data on the identities and the respective entitlements assigned to each identity as utilized in an enterprise computer environment may be obtained by an identity management system. Using the identity and

entitlement data, then, a network identity graph may be constructed, where the nodes of the graph correspond to, and represent, each of the identities or entitlements. Each edge (or relationship) of the graph may join two nodes of the graph and be associated with a similarity weight representing a degree of similarity between the identities or entitlements of the respective nodes. The identity graph may then be pruned to remove weak edges (e.g., those edges whose similarity weight may fall below a pruning threshold). The pruned identity graph can then be clustered into peer groups of identities or entitlement groups (e.g., using a graph based community detection algorithm). These peer groups of identities (or entitlements) can then be stored (e.g., separately or in the identity graph) and used by the identity management system. For example, a visual representation of the graph may be presented to a user of the identity management to assist in compliance or certification assessments or evaluation of the identities and entitlements as currently used by the enterprise.

In certain embodiments, the clustering of identities or entitlements may be optimized based on a peer group assessment metric, such as, for example, graph modularity determined based on the identity graph or the determined peer groups. For instance, in one embodiment, if a peer group assessment metric is below (or above) a quality threshold a feedback loop may be instituted whereby the pruning threshold is adjusted by some amount (up or down) and the originally determined identity graph is pruned based on the adjusted pruning threshold (or the previously pruned identity graph may be further pruned). This newly pruned identity graph can then be clustered into new peer groups of identities or entitlements and a peer group assessment metric determined based on the newly pruned identity graph or the newly determined peer groups. If this new peer assessment metric is now above (or below) the quality threshold the feedback loop may stop and these peer groups of identities or entitlements can then be stored (e.g., separately or in the identity graph) and used by the identity management system.

Otherwise, the feedback loop may continue by again adjusting the pruning threshold further (e.g., further up or further down relative to the previous iteration of the feedback loop), re-pruning the identity graph based on the adjusted pruning threshold, clustering this newly pruned graph, determining another peer group assessment metric and comparing this metric to the quality threshold. In this manner, the feedback loop of adjustment of the pruning threshold, re-pruning the graph, re-clustering the identity graph into peer groups may be repeated until the peer group assessment metric reaches a desired threshold. Moreover, by tailoring the peer group assessment metric and quality threshold to include or reflect domain or enterprise specific criteria, the clustering results (e.g., the peer groups of identities or entitlements resulting from the clustering) may more accurately reflect particular requirements or the needs of a particular enterprise or be better tailored to a particular use.

Embodiments may thus provide a number of advantages including allowing more intuitive access to the data (e.g., via graph database querying, or via graph visualization techniques), which may, in turn, yield deeper and more relevant insights for users of identity management systems. Moreover, embodiments as disclosed may offer the technological improvement of reducing the computational burden and memory requirements of systems implementing these embodiments through the improved data structures and the graph processing and analysis implemented by such embodiments. Accordingly, embodiments may improve the performance and responsiveness of identity management systems that utilize such embodiments. Likewise, embodiments may be dynamic with respect to time, allowing the development update processes using deltas between snapshots of data collection, bringing down operational costs and improving the performance and robustness of embodiments. Moreover, the graph format used by certain embodiments, allows the translation of domain and enterprise specific concepts, phenomena, and issues into tangible, quantifiable, and verifiable hypotheses which may be examine or validate with graph based algorithms. Accordingly, embodiments may be especially useful in assessing risk and in compliance with security policies or the like.

Turning first to FIG. **1**, then, a distributed networked computer environment including one embodiment of an identity management system is depicted. Here, the networked computer environment may include an enterprise computing environment **100**. Enterprise environment **100** includes a number of computing devices or applications that may be coupled over a computer network **102** or combination of computer networks, such as the Internet, an intranet, an internet, a Wide Area Network (WAN), a Local Area Network (LAN), a cellular network, a wireless or wired network, or another type of network. Enterprise environment **100** may thus include a number of resources, various resource groups and users associated with an enterprise (for purposes of this disclosure any for profit or non-profit entity or organization). Users may have various roles, job functions, responsibilities, etc. to perform within various processes or tasks associated with enterprise environment **100**. Users can include employees, supervisors, managers, IT personnel, vendors, suppliers, customers, robotic or application based users, etc. associated with enterprise **100**.

Users may access resources of the enterprise environment **100** to perform functions associated with their jobs, obtain information about enterprise **100** and its products, services, and resources, enter or manipulate information regarding the same, monitor activity in enterprise **100**, order supplies and services for enterprise **100**, manage inventory, generate financial analyses and reports, or generally to perform any task, activity or process related to the enterprise **100**. Thus, to accomplish their responsibilities, users may have entitlements to access resources of the enterprise environment **100**. These entitlements may give rise to risk of negligent or malicious use of resources.

Specifically, to accomplish different functions, different users may have differing access entitlements to differing resources. Some access entitlements may allow particular users to obtain, enter, manipulate, etc. information in resources which may be relatively innocuous. Some access entitlements may allow particular users to manipulate information in resources of the enterprise **100** which might be relatively sensitive. Some sensitive information can include human resource files, financial records, marketing plans, intellectual property files, etc. Access to sensitive information can allow negligent or malicious activities to harm the enterprise itself. Access risks can thus result from a user having entitlements with which the user can access resources that the particular user should not have access to; or for other reasons. Access risks can also arise from roles in enterprise environment **100** which may shift, change, evolve, etc. leaving entitlements non optimally distributed among various users.

To assist in managing the entitlements assigned to various users and more generally in managing and assessing access risks in enterprise environment **100**, an identity management system **150** may be employed. Such an identity management

system **150** may allow an administrative or other type of user to define one or more identities, one or more entitlements, or one or more roles, and associate defined identities with entitlements using, for example, an administrator interface **152**. The assignment may occur, for example, by directly assigning an entitlement to an identity, or by assigning a role to an identity whereby the collection of entitlements comprising the role are thus associated with the identity. Examples of such identity management systems are Sailpoint's IdentityIQ and IdentityNow products. Note here, that while the identity management system **150** has been depicted in the diagram as separate and distinct from the enterprise environment **100** and coupled to enterprise environment **100** over a computer network **104** (which may the same as, or different than, network **102**), it will be realized that such an identity management system **150** may be deployed as part of the enterprise environment **100**, remotely from the enterprise environment, as a cloud based application or set of services, or in another configuration.

An identity may thus be almost physical or virtual thing, place, person or other item that an enterprise would like to define. For example, an identity may be a capacity, groups, processes, physical locations, individual users or humans or almost any other physical or virtual entity, place, person or other item. An entitlement may be an item (e.g., token) that upon granting to a user will allow the user to acquire a certain account or privileged access level that enables the user to perform a certain function within the distributed networked enterprise computer environment **100**. Thought of another way, an entitlement may be a specific permission granted within a computer system, such as access to a particular building (based on a user's key badge), access to files and folders, or access to certain parts of websites. Entitlements may also define the actions a user can take against the items they have access to, including, for example, accessing computing systems, applications, file systems, particular data or data items, networks, subnetworks or network locations, etc. Each of these identities may therefore be assigned zero or more entitlements with respect to the distributed networked computer environments.

To facilitate the assignment of these entitlements, enterprises may also be provided with the ability to define roles through the identity management system **150**. A role within the context of the identity management system **150** may be a collection of entitlements. These roles may be assigned a name or identifiers (e.g., manager, engineer_level_2, team leader) by an enterprise that designate the type of user or identity that should be assigned such a role. By assigning a role to an identity using the identity management system **150**, the identity may be assigned the corresponding collection of entitlements associated with the assigned role.

The identity management system **150** may thus store identity management data **154**. The identity management data **154** stored may include a set of entries, each entry corresponding to and including an identity (e.g., alphanumerical identifiers for identities) as defined and managed by the identity management system, a list or vector of entitlements or roles assigned to that identity by the identity management system, and a time stamp at which the identity management data was collected from the identity management system. Other data could also be associated with each identity, including data that may be provided from other systems such as a title, location or department associated with the identity. The set of entries may also include entries corresponding to roles, where each entry for a role may include the role identifier (e.g., alphanumerical identifier or name for the role) and a list or vector of the entitlements associated with each role. Other data could also be associated with each role, such as a title, location or department associated with the role.

Collectors **156** of the identity management system **150** may thus request or otherwise obtain data from various touchpoint systems within enterprise environment **100**. These touchpoint systems may include, for example Active Directory systems, Java Database Connectors within the enterprise **100**, Microsoft SQL servers, Azure Active Directory servers, OpenLDAP servers, Oracle Databases, SalesForce applications, ServiceNow applications, SAP applications or Google GSuite.

Accordingly, the collectors **156** of the identity management system **150** may obtain or collect event data from various systems within the enterprise environment **100** and process the event data to associate the event data with the identities defined in the identity management data **154** to evaluate or analyze these events or other data in an identity management context. A user may interact with the identity management system **150** through a user interface **158** to access or manipulate data on identities, roles, entitlements, events or generally preform identity management with respect to enterprise environment **100**.

As part of a robust identity management system, it is desirable to analyze the identity management data **154** associated with an enterprise **100**. Specifically, it is desirable to group or cluster the identities or entitlements of an enterprise **100** into peer groups such that, for example, the identities in a peer group are similar with respect to the set of entitlements assigned to the identities of that group (e.g., relative to other identities or other groups) or, to determine peer groups of entitlements such that entitlement patterns and assignment may be determined and role mining performed.

Peer grouping of the identities within an enterprise (or viewing the peer groups of identities) may allow, for example, an auditor other person performing a compliance analysis or evaluation to quantitatively and qualitatively assess the effectiveness of any applicable pre-existing polices, or lack thereof, and how strictly they are enforced. Similarly, peer grouping of entitlements may allow roles to be determined from such entitlement groups and outlier entitlements to be identified. This information may, in turn, be utilized to redefine or govern existing roles as defined in the identity management system **150** and allow users of the identity management system **150** greater visibility into the roles of the enterprise **100**.

Accordingly, an identity management system **160** may include a harvester **162** and a graph generator **164**. The harvester **162** may obtain identity management data from one or more identity management systems **150** associated with enterprise **100**. The identity management data may be obtained, for example, as part of a regular collection or harvesting process performed at some regular interval by connecting to, and requesting the identity management data from, the identity management system **150**. The identity management data stored may thus include a set of entries, each entry corresponding to and including an identity as defined and managed by the identity management system, a list or vector of entitlements or roles assigned to that identity by the identity management system, and a time stamp at which the identity management data was collected from the identity management system **150**. The identity management data may also include a set of entries for roles, each entry corresponding to and including a role as defined and managed by the identity management system **150** and a list or vector of entitlements assigned to that role by the identity

management system **150**, and a time stamp at which that identity management data was collected from the identity management system **150**.

Graph generator **164** may generate a peer grouped identity graph from the obtained identity management data. Specifically, in one embodiment, a property (identity) graph may be generated from the identity management data obtained from the enterprise. Each of the identities and entitlements from the most recently obtained identity management data may be determined and a node of the graph created for each identity and entitlement. An edge is constructed between every pair of nodes (e.g., identities) that shares at least one entitlement and between every pair of nodes (e.g., entitlements) that shares at least one identity. Each edge of the graph may also be associated with a similarity weight representing a degree of similarity between the identities of the respective nodes joined by that edge, or between the entitlements of the respective nodes joined by that edge. It will be noted here that while a similarity weight may be utilized on edges between both identity nodes and entitlement nodes, the similarity weight type, determination and value may be determined differently based upon the respective type of node(s) being joined that weighted edge. Accordingly, the obtained identity management data may be represented by an identity graph (e.g., per enterprise) and stored in graph data store **166**.

Once the identity graph is generated by the graph generator **164**, the graph may then be pruned to remove edges based on their weighting. Again, the pruning of edges between identity nodes and entitlements nodes may be accomplished in the same, or a different manner. For example, a pruning threshold utilized to prune edges between identity nodes may be different than a pruning threshold utilized to prune edges between entitlement nodes as well as across customers.

The pruned identity graph can then be used to cluster the identities into peer groups of identities or to cluster the entitlements into peer groups of entitlements. This clustering may be accomplished, for example, a community-detection algorithm. This clustering result may also be optimized by the graph generator **164** through the use of a feedback loop to optimize the pruning of the edges until a desired metric for assessing the quality of the peer groups generated exceeds a desired threshold or satisfies certain (e.g., optimization or other) criteria. It will be noted here as well, that while the peer grouping of both identities or entitlements may be determined in embodiments, the peer grouping may be accomplished in the same or different manners for identities and entitlements in different embodiments. For example, the community detection, optimization, feedback loop or quality assessment metric may all be the same or different when clustering the identity or entitlements of the entitlement graph. It will also be noted here, that while identities and entitlements are discussed herein as examples of identity management artifacts that are represented as nodes in the graph, as discussed above, other identity management artifacts (e.g., roles, groups, etc.) may also be represented as nodes in the identity graph, and may be similar clustered or grouped into peer groups.

More generally, then, the pruning and clustering of the identity nodes of the identity graph may be performed separately from the pruning and clustering of the entitlement nodes of the identity graph. Accordingly, the property graph may comprise at least two subgraphs, the identities subgraph comprising at least the identity nodes and edges between these identity nodes and the entitlement subgraph comprising at least the entitlement nodes and edges between those entitlement nodes. Once the peer groups of identities or entitlements are determined, the peer groups can then be stored (e.g., separately or in the property graph itself) and used by the identity management system **160**. For example, each peer group of identities (also referred to herein as an identity group) may be assigned a peer group identifier and the peer group identifier associated with each identity assigned to the peer group by storing the peer group identifier in association with the node in the graph representing that identity. Similarly, each peer group of entitlements (e.g., also referred to herein as an entitlement group) may be assigned a peer group identifier and the peer group identifier associated with each entitlement assigned to the peer group by storing the peer group identifier in association with the node in the graph representing that entitlement.

An interface **168** of the identity management system **160** may use the identity graph in the graph data store **166** or associated peer groups to present one or more interface which may be used for risk assessment, as will be discussed. For example, an interface **168** may present a visual representation of the graph, the identities, entitlements, or the peer groups in the identity graph to a user of the identity management system **160** associated with enterprise **100** to assist in compliance or certification assessments or evaluation of the identities, entitlements or roles as currently used by the enterprise (e.g., as represented in identity management data **154** of identity management system **150**).

Before moving on, it will be noted here that while identity management system **160** and identity management system **150** have been depicted separately for purposes of explanation and illustration, it will be apparent that the functionality of identity management systems **150**, **160** may be combined into a single or a plurality of identity management system as is desired for a particular embodiment and the depiction and separation of the identity management systems and their respective functionality has been depicted separately solely for purposes of ease of depiction and description.

Turning now to FIG. **2**, a flow diagram for one embodiment of a method for determining peer groups of identities using a graph database is depicted. Embodiments of such a method may be employed by graph generators of identity management systems to generate identity graphs and associated peer groups from identity management data, as discussed above. It will be noted here, that while this embodiment is described in association with the determination of peer groups of identities in the identity graph, similar embodiments may be applied to entitlement nodes and associated similarity relationships of an identity graph to determine peer groups of entitlements in such an identity graph.

Initially, at step **210**, identity management data may be obtained. As discussed, in one embodiment, this identity management data may be obtained from one or more identity management systems that are deployed in association with an enterprise's distributed computing environment. Thus, the identity management data may be obtained, for example, as part of a regular collection or harvesting process performed at some regular interval by connecting to, requesting the identity management data from, an identity management system. The identity management data may also be obtained on a one-time or user initiated basis.

As will be understood, the gathering of identity management data and determination of peer groups can be implemented on a regular, semi-regular or repeated basis, and thus may be implemented dynamically in time. Accordingly, as the data is obtained, it may be stored as a time-stamped snapshot. The identity management data stored may thus

include a set of entries, each entry corresponding to and including an identity (e.g., alphanumerical identifiers for identities) as defined and managed by the identity management system, a list or vector of entitlements assigned to that identity by the identity management system, and a time stamp at which the identity management data was collected from the identity management system. Other data could also be associated with each identity, including data that may be provided from an identity management system such as a title, location or department associated with the identity. The collection of entries or identities associated with the same times stamp can thus be thought of as a snapshot from that time of the identities and entitlements of the enterprise computing environment as management by the identity management system.

As an example of identity management data that may be obtained from an identity management system, the following is one example of a Javascript Object Notation (JSON) object that may relate to an identity:


```
{
  "attributes": {
    "Department": "Finance",
    "costcenter": "[R01e, L03]",
    "displayName": "Catherine Simmons",
    "email": "Catherine.Simmons@demoexample.com",
    "empId": "1b2c3d",
    "firstname": "Catherine",
    "inactive": "false",
    "jobtitle": "Treasury Analyst",
    "lastname": "Simmons",
    "location": "London",
    "manager": "Amanda.Ross",
    "region": "Europe",
    "riskScore": 528,
    "startDate": "12/31/2016 00:00:00AM UTC",
    "nativeIdentity_source_2": "source_2",
    "awesome_attribute_source_1": "source_1",
    "twin_attribute_a" : "twin a",
    "twin_attribute_b" : "twin b",
    "twin_attribute_c" : "twin c"
  },
  "id": "2c9084ee5a8de328015a8de370100082",
  "integration_id": "iiq",
  "customer_id": "ida-bali",
  "meta": {
    "created": "2017-03-02T07:19:37.233Z",
    "modified": "2017-03-02T07:24:12.024Z"
  },
  "name": "Catherine. Simmons",
  "refs": {
    "accounts": {
      "id": [
          "2c9084ee5a8de328015a8de370110083"
        ],
        "type": "account"
      },
      "entitlements": {
      "id": [
        "2c9084ee5a8de328015a8de449060e54",
        "2c9084ee5a8de328015a8de449060e55"
      ],
      "type": "entitlement"
    },
    "manager": {
      "id": [
      "2c9084ee5a8de022015a8de0c52b031d"
    ],
    "type": "identity"
    }
  },
  "type": "identity"
}
```


As another example of identity management data that may be obtained from an identity management system, the following is one example of a JSON object that may relate to an entitlement:


```
{
  "integration_id": "bd992e37-bbe7-45ae-bbbf-
    c97a59194cbc",
  "refs": {
    "application": {
      "id": [
        "2c948083616ca13a01616ca1d4aa0301"
      ],
        "type": "application"
      }
    },
    "meta": {
      "created": "2018-02-06T19:40:08.005Z",
      "modified": "2018-02-06T19:40:08.018Z"
    },
    "name": "Domain Administrators",
    "attributes": {
        "description": "Domain Administrators group
        on Active Directory",
        "attribute": "memberOf",
        "aggregated": true,
        "requestable": true,
        "type": "group",
        "value": "cn=Domain Administrators,
        dc=domain,dc=local"
      },
      "id": "2c948083616ca13a01616ca1f1c50377",
      "type": "entitlement",
      "customer_id": "3a60b474-4f43-4523-83d1-
        eb0fd571828f"
    }
```


As another example of identity management data that may be obtained from an identity management system, the following is one example of a JSON object that may relate to a role:


```
{
"id": "id",
"name": "name",
"description":
"description",
"modified": "2018-09-
07T17:49:33.667Z", "created":
"2018-09-07T17:49:33.667Z",
"enabled": true,
"requestable": true,
"tags": [
{
  "id": "2c9084ee5a8ad545345345a8de370110083"
    "name": "SOD-SOX",
      "type": "TAG"
  },
{
  "id": "2c9084ee5a8ad545345345a8de370122093"
    "name": "PrivilegedAccess",
      "type": "TAG"
  },
]
"accessProfiles":
[
  {
```

"id": "accessProfileId",
"name": "accessProfileName"
}
],
"accessProfileCount":
1, "owner": {
"name": "displayName",
"id": "ownerId"
},
"synced": "2018-09-07T17:49:33.667Z"
}

At step **220** an identity graph may be generated from the identity management data obtained from the enterprise. Specifically, each of the identities and entitlements from the most recent snapshot of identity management data may be obtained and a node of the graph created for each identity and entitlement. An edge is constructed between every pair of identity nodes (e.g., identities) that shares at least one entitlement (e.g., an edge connects two identity nodes if and only if they have at least one entitlement in common). An edge may also be constructed between every pair of entitlement nodes (e.g., entitlements) that shares at least one identity (e.g., an edge connects two entitlement nodes if and only if they have at least one identity in common).

Each edge of the graph joining identity nodes or entitlement nodes may be associated with a similarity weight representing a degree of similarity between the identities or entitlements of the respective nodes joined by that edge. For identity nodes, the similarity weight of an edge joining the two identity nodes may be generated based on the number of entitlements shared between the two joined nodes. As but one example, the similarity weight could be based on a count of the similarity (e.g., overlap or intersection of entitlements) between the two identities divided by the union of entitlements. Similarly, for identity nodes, the similarity weight of an edge joining the two entitlement nodes may be generated based on the number of identities shared between the two joined nodes. As but one example, the similarity weight could be based on a count of the similarity (e.g., overlap or intersection of identities) between the two entitlements divided by the union of identities. For instance the similarity could be defined as the ratio between a number of identities having both entitlements joined by the edge to the number of identities that have either one (e.g., including both) of the two entitlements.

In one embodiment, the edges are weighted via a proper similarity function (e.g., Jaccard similarity). In one embodiment, a dissimilarity measure, of entitlement or identity binary vectors, d, may be chosen, then the induced similarity, 1−d(x,y), may be used to assign a similarity weight to the edge joining the nodes, x,y. Other methods for determining a similarity weight between two nodes are possible and are fully contemplated herein. Moreover, it will be noted here that while a similarity weight may be utilized on edges between both identity nodes and entitlement nodes, the similarity weight type, determination and value may be determined differently based upon the respective type of node(s) being joined that weighted edge.

In one specific, embodiment, a symmetric matrix for identities (e.g., an identity adjacency matrix) may be determined with each of the identities along each axis of the matrix. The diagonal of the matrix may be all Os while the rest of values are the similarity weights determined between the two (identity) nodes on the axes corresponding to the value. In this manner, this symmetric matrix may be provided to a graph constructor which translates the identities on the axes and the similarity values of the matrix into graph store commands to construct the identity graph. Similarly, a symmetric matrix for entitlements (e.g., an entitlement adjacency matrix) may be determined with each of the entitlements along each axis of the matrix. The diagonal of the matrix may be all Os while the rest of values are the similarity weights determined between the two (entitlement) nodes on the axes corresponding to the value. In this manner, this symmetric matrix may be provided to a graph constructor which translates the entitlement on the axes and the similarity values of the matrix into graph store commands to construct the identity graph.

Accordingly, the identity management data may be faithfully represented by a graph, with k types of entities (nodes/vertices, e.g., identity-id, title, location, entitlement, etc.) and stored in a graph data store. It will be noted that graph data store **132** may be stored in any suitable format and according to any suitable storage, including, for example, a graph store such a Neo4j, a triple store, a relational database, etc. Access and queries to this graph data store may thus be accomplished using an associated access or query language (e.g., such as Cypher in the case where the Neo4j graph store is utilized).

Once the identity graph is generated, the graph may then be pruned at step **230**. Here, the identity graph may be pruned to remove weak edges (e.g., those edges whose similarity weight may fall below a pruning threshold). The pruning of the graph is associated with the locality aspect of identity governance, where an identity's access entitlements should not be directly impacted, if at all, by another identity with strongly dissimilar entitlement pattern (e.g., a weak connecting edge) or that determined should be based on strong commonality or popularity of entitlements within an identity grouping. Accordingly, the removal of such edges may not dramatically alter the global topology of the identity graph. An initial pruning threshold may be initially set or determined (e.g., as 50% similarity or the like) and may be substantially optimized or otherwise adjusted at a later point. As another example, a histogram of similarity weights may be constructed and a similarity weight corresponding to a gap in the similarity weights of the histogram may be chosen as an initial pruning threshold. Again, the pruning of edges between identity nodes and entitlements nodes may be accomplished in the same, or a different manner. For example, the pruning threshold utilized to prune edges between identity nodes may be different than a pruning threshold utilized to prune edges between entitlement nodes.

The pruned identity graph can then be used to cluster the identities or entitlements into peer groups of identities or peer groups of entitlements at step **240**. Within this graph approach, a representation of a peer group of identities could be represented by a maximal clique, where every identity is strongly connected (e.g., similar) to every other identity within the identity peer group, and consequently, members of the clique all share a relatively large, and hence dominant, common core of entitlements. A representation of an entitlement peer group could be represented by a maximal clique, where every entitlement is strongly connected (e.g., similar) to every other entitlement within the peer group, and consequently, members of the clique all share a relatively large, and hence dominant, common core of identities. The problem of finding all maximal cliques of a graph may, however, be a memory and computationally intensive problem. Most clique related problems in graph theory are hard and some of them are even NP-complete, requiring exponential time to finish as graphs with exponentially many maximal cliques may exist.

Accordingly, in one embodiment a community-detection algorithm may be utilized for peer grouping the identities or entitlements of the identity graph to speed the determination of the peer groups, reduce computational overhead and conserve memory, among other advantages. A plethora of applicable and performant community-detection and graph clustering algorithms may be utilized according to certain embodiments. Some of these algorithms are specifically targeted to large graphs, which can be loosely described as graphs with at least tens or hundreds (or more) of thousands of nodes and millions of edges. Such graph community-detection algorithms may include, for example, Louvain, Leiden, Fast-greedy, Label Propagation or Stochastic Block Modeling. Other graph community detection algorithms may be utilized and are fully contemplated herein.

In certain embodiments, a clustering result may be optimized through the use of a feedback loop, as discussed below. As such, in one embodiment it may be desirable to utilize a community-detection algorithm for determination of the peer groups that may provide allow a straightforward determination of a peer group assessment metric for a quality assessment of determined peer groups or the identity graph. Accordingly, a community-detection algorithm that may be based on, or allow a determination of, a graph based metric (e.g., modularity, evolving topology, connected components, centrality measures (e.g., betweenness, closeness, community overlap measures such as NMI or Omega indices)) that may be used as a peer group assessment metric may be utilized.

Specifically, in one embodiment, the Louvain algorithm may be utilized as a community-detection algorithm and modularity may be used as a peer assessment metric. The Louvain algorithm may not only be a scalable algorithm that can handle, and be efficient on, large graphs; but additionally the Louvain algorithm may be based on modularity or be modularity optimized. Modularity is a scalar that can be determined for a graph or groups or subgraphs thereof. This modularity reflects a likelihood of the clusters generated (e.g., by the algorithm) to not have been generated by random chance. A high modularity value, (e.g., positive and away from 0) may indicate that the clustering result is unlikely to be a product of chance. This modularity can thus be used as a peer group assessment metric.

Moreover, in addition to the application of a peer group assessment metric to optimize the peer groups or identity graphs determined using such community-detection algorithms, an identity management system may employ alerts based these peer group assessment metrics. For example, an alert to a user may be based on an alert threshold (e.g., if the peer group assessment metric drops below or above a certain threshold) or if any changes over a certain threshold occur with respect to the peer group assessment metric. For example, setting an empirical low threshold for modularity, with combined user alerts, could serve as a warning for deteriorating quality of peer groups or the identity graph. This could be due to input data has been corrupted at some point in pipeline, or in other cases, that the access entitlement process for the particular enterprise is extremely lacking due discipline. Regardless of the underlying cause, such an early warning system may be valuable to stop the propagation of questionable data quality in the peer group assessment and determination process and more generally to identity management goals within the enterprise.

In many cases, the community-detection or other clustering algorithm utilized in an embodiment may fall under the umbrella of what are usually termed unsupervised machine-learning. Results of these types of unsupervised learning algorithms may leave some room for interpretation, and do not, necessarily or inherently, provide outputs that are optimized when the domain or context in which they are being applied are taken into account. Consequently, to mitigate some of these issues and to optimize the use of the peer groups and identity graphs in an identity governance context, embodiments of identity management systems employing such peer groups of identities or entitlements using an identity graph may allow some degree of user configuration, where at a least a portion of the user configuration may be applied in the graph determination, peer-grouping or optimization of such peer group determination.

This configurability may allow the user of an identity management system to, for example, impose some constraints or set up certain configuration parameters for the community-detection (or other peer grouping) algorithm in order to enhance the clustering results for a particular use-case or application. A few non-exhaustive examples of user configuration are thus presented. A user may have a strongly defined concept of what constitutes a 'peer'. This may entail that the user's specification of what continues a peer may be used to derive a pruning threshold with statistical methods (e.g., rather than relying on modularity).

As another example of configurability, a user may elect to opt for a hierarchical clustering output, or that peer groups should have certain average size, which may entail to allowing for several consecutive iterations of the community-detection algorithm to be performed (as will be explained in more detail herein). A user may also elect to run the peer grouping per certain portions of the identities or entitlements, versus running it for all identities or entitlements. The filtered population of identities or entitlements may be specified in terms of geographic location, business role, business unit, etc. Similarly, a user may elect to filter the outputs of the community-detection algorithm in terms of certain identity or entitlement attributes, e.g., identity role, identity title, identity location, etc. The results might then be quantitatively and qualitatively contrasted against existing governance policies to measure, assess and certify compliance with these policies.

Generally then, a user may elect to utilize the peer grouping feature in combination with other tools of identity governance, in order to gain more insight into the quality of identity governance policy enforcement within the business. This entails that peer grouping should be configurable and flexible enough to allow it to be paired with other (e.g., third-party) identity management tools. Accordingly, certain restrictions may be imposed on the identity graph's or peer group's size, format, level of detail, etc.

In any event, once the peer groups of identities or entitlements of the pruned identity graph are used to cluster the identities into peer groups of identities at step **240** the determined peer groups can then be stored (e.g., separately or in the identity graph itself) and used by the identity management system. For example, each peer group (e.g., or identities or entitlements) may be assigned a peer group identifier and the peer group identifier associated with each identity assigned to the peer group by storing the peer group identifier in association with the node in the graph representing that identity or entitlement.

As an example of use a visual representation of the graph, the identities, entitlements or the peer groups in the identity graph may be presented to a user of the identity management to assist in compliance or certification assessments or evaluation of the identities and entitlements as currently used by the enterprise. In principle, strictly enforced pre-existing governance policies should ensure that identities with

strongly similar access privileges are strongly similar (e.g., are in the same peer group). The presentation of such peer groups may thus, for example, allow an auditor or compliance assessor to quantitatively and qualitatively assess the effectiveness of any applicable pre-existing polices, or lack thereof, and how strictly they are enforced.

During such collection, graph determination and peer grouping steps, in certain embodiments, a number of efficiencies may be implemented to speed the collection process, reduce the amount data that must be stored and to reduce the computer processing overhead and computing cycles associated with such data collection, graph determination and peer grouping of such data. Specifically, in one embodiment, a delta change assessment may be performed when identity management data is collected or peer groups are determined in a current time period. More specifically, if identity management data was collected in a previous time period, or a previous peer grouping has been performed on identities or entitlements of a previously created identity graph, an assessment can be made (e.g., by a data querying script or process) of the difference (or delta) between the set of identities or entitlements corresponding to the most recent previous snapshot and the set of identities or entitlements obtained in the current time period. This assessment may comprise a determination of how many changes to the identities, associated entitlements or other attributes have occurred between the time of the previous snapshot and the current snapshot (e.g., the most recently identity management data collected in the current time period).

An assessment may also be made of the difference between the peer groups determined from the most recent previous snapshot and the peer groups obtained in the current time period. This assessment may comprise a determination of how many identities or entitlements are associated with different peer groups (e.g., relative to the peer grouping of identities or entitlements determined from the previous most recent snapshot), changes to the identities or entitlements or how many new identities are associated with an established (or new) peer group.

If there are no determined changes, or the changes are below some threshold number, or are few, local, or insignificant to a large majority of existing peer groups, then no action is needed other than updating the affected identities or entitlements in the data of the previous snapshot or the identity graph. New entries in the entries comprising the current snapshot of identities or entitlements may be created for any newly identified identities or entitlements. Additionally, nodes in the graph corresponding to new identities or entitlements can be appended to an appropriate peer group based on how similar this new identity to existing peer groups, (e.g., assign the new identity the peer group of the same department/title).

If the differences (e.g., number of changes, new identities, different peer group assignments, etc.) are non-trivial, affecting a multitude of identities across peer groups, then a new peer grouping process may occur on the newly refreshed data. In such case, a detection algorithm may be used to evolve, and persist, previously determined peer groups into their recent counterparts. This can be done by monitoring certain 'marker' identities, e.g., influencers, or identities with high centrality values and/or high degree of connections, in both versions of peer groups. Utilizing a majority vote approach, it can be determined how previous peer groups evolve into newer ones. Expected updated versions of the previous peer group, include splitting, merging, growth, shrinkage. Newer split peer groups may, for example inherit the 'old' peer group identifiers.

Embodiments of such a delta detection and updating mechanisms may have the further advantage of allowing the quality and stability of each peer group to be monitored by an identity management system via tracking the peer groups or identity graph, the changes thereto, or their evolution over time. By actively monitoring and assessing the degree of these changes between two or more consecutive versions of a peer group or identity graph, deteriorating quality issues may be detected as they arise or manifest in the identity graph or peer groups determined therefrom. Similarly, using the identity graphs, peer groups or peer group assessment metrics determined therefrom, a graph evolution model may be built in certain embodiments, (e.g., based on epidemiology susceptible, infected and recovered type models). Comparing the observed evolution of identities, entitlements or peer groups versus theoretical predictions may provide another tool to warn users of an identity management system against rapid or extreme changes that may negatively impact the quality of peer groups or identity management more generally.

Again, once the peer groups of identities or entitlements are determined from the pruned identity graph and stored (at step **240**), a peer group assessment metric may be determined based on the identity graph or the determined peer groups at step **250**. As discussed, this peer group assessment metric may be determined separately based on the peer groups or identity graph determined, or may be metric utilized by a community-detection algorithm, such that the peer group assessment metric may be determined as part of the peer group determination process. In certain embodiments then, the application of a community-detection algorithm may result in such a peer group assessment metric (e.g., modularity, evolving topology, connected components, centrality measures e.g., betweenness, closeness, community overlap measures (e.g., NMI, Omega indices)) that may be used as a peer group assessment metric may be utilized.

For example, as discussed above the Louvain algorithm may be a graph-based modularity optimized community-detection algorithm. Thus, a modularity associated with the determined peer groups may result from the determination of the peer group using the Louvain algorithm. Modularity is a scalar that can be determined for a graph or groups or subgraphs thereof and reflects a likelihood of the clusters generated (e.g., by the algorithm) to not have been generated by random chance. A high modularity value, (e.g., positive and away from 0) may indicate that the clustering result is unlikely to be a product of chance. This modularity can be used as a peer group assessment metric in one embodiment.

Accordingly, in certain embodiments, the clustering of identities or entitlements into peer groups may be optimized based on this peer group assessment metric. Specifically, a feedback loop may be utilized to determine the optimal pruning threshold. The optimization loop may serve to substantially increase or maximize the quality of the graph clustering, with respect to certain proper metrics (e.g., graph modularity or other peer group assessment metric). Additional domain-specific, per enterprise, criteria may be utilized in this step in certain embodiments in order to render clustering results that accurately reflect certain requirements to better serve a particular enterprise or use of the peer groups or identity graph.

For instance, in one embodiment if the peer group assessment metric is above (or below) a quality threshold at step **260** the determination of peer groups of identities or entitlements for the obtained in the current snapshot may end at step **262**. The determined peer groups of identities or entitle-

ments can then be stored (e.g., separately or in the identity graph) and used by the identity management system.

However, if the peer group assessment metric is below (or above) a quality threshold at step **260** a feedback loop may be instituted whereby the pruning threshold is adjusted by some amount at step **270** (up or down) and the originally determined identity graph is again pruned based on the adjusted pruning threshold (or the previously pruned identity graph may be further pruned) at step **230**. The adjustment of the pruning threshold may be based on a wide variety of criteria in various embodiments and may be adjust be a fixed or differing amount in every iteration through the feedback loop. Additionally, in some embodiments, various machine learning techniques (e.g., unsupervised machine learning techniques such as k-means, method of moments, neural networks, etc.) may be used to determine an amount to adjust the pruning threshold or a value for the adjusted pruning threshold). This newly pruned identity graph can then be clustered into new peer groups of identities or entitlements at step **240** and a peer group assessment metric determined at step **250** based on the newly pruned identity graph or the newly determined peer groups.

If this new peer assessment metric is now above (or below) the quality threshold at step **260** the feedback loop may be stopped and the determination of peer groups of identities or entitlements for the data obtained in the current snapshot may end at step **262**. These peer groups of identities or entitlements can then be stored (e.g., separately or in the identity graph) and used by the identity management system.

Otherwise, the feedback loop may continue by again adjusting the pruning threshold further at step **270** (e.g., further up or further down relative to the previous iteration of the feedback loop), re-pruning the identity graph based on the adjusted pruning threshold at step **230**, clustering this newly pruned graph at step **240**, determining another peer group assessment metric at step **250** and comparing this metric to the quality threshold at step **260**. In this manner, the feedback loop of adjustment of the pruning threshold, re-pruning the graph and re-clustering the identity graph into peer groups may be repeated until the peer group assessment metric reaches a desired threshold. Moreover, by tailoring the peer group assessment metric and quality threshold to include or reflect domain or enterprise specific criteria (e.g., which may be specified by a user of the identity management system), the clustering results (e.g., the peer groups resulting from the clustering) may more accurately reflect particular requirements or the needs of a particular enterprise or be better tailored to a particular use.

Once the feedback loop is ended (step **262**) the determined peer groups of identities or entitlements can then be stored (e.g., separately or in the identity graph) and used by the identity management system. For example, a visual representation of the graph may be presented to a user of the identity management to assist in compliance or certification assessments or evaluation of the identities and entitlements as currently used by the enterprise.

It will be noted here as well, that while the peer grouping of both identities or entitlements may be determined in embodiments, the peer grouping may be accomplished in the same or different manners for identities and entitlements in different embodiments. For example, the community detection, optimization, feedback loop or quality assessment metric (e.g., steps **230**, **240**, **250**, **260** and **270**) may all be performed the same or differently when clustering the identity or entitlements of the entitlement graph. More generally, then, the pruning and clustering of the identity nodes of the identity graph may be performed separately from the pruning and clustering of the entitlement nodes of the identity graph. In certain embodiments, for example, the pruning and clustering (e.g., steps **230**, **240**, **250**, **260** and **270**) of the identity nodes of the identity graph may be performed as a separate process from the pruning and clustering (e.g., steps **230**, **240**, **250**, **260** and **270**) of the entitlement nodes of the identity graph. Accordingly, the identity graph may be comprised of at least two subgraphs, the identities subgraph comprising at least the identity nodes and edges between these identity nodes and the entitlement subgraph comprising at least the entitlement nodes and edges between those entitlement nodes.

It may now be helpful to look at such visual depictions and presentations of identity graphs or interfaces that may be created or presented based on such identity graphs. It will be apparent that these depictions and interfaces are but example of depictions and interfaces that may presented or utilized, and that almost any type of presentation, depiction or interface based on the identities, entitlements, peer groups or other associated data discussed may be utilized in association with the embodiments of identity management systems disclosed herein.

As discussed, embodiments of the identity management systems as disclosed may create, maintain or utilize identity graphs. These identity graphs may include a graph comprised of nodes and edges, where the nodes may include identity management nodes representing, for example, an identity, entitlement or peer group, and the edges may include relationships between these identity management nodes. The relationships represented by the edges of the identity graph may be assigned weights or scores indicating a degree of similarity between the nodes related by a relationship, including, for example, the similarity between two nodes representing an identity or two nodes representing an entitlement, as discussed. Additionally, the relationships may be directional, such that they may be traversed only in a single direction, or have different weightings depending on the direction in which the relationship is traversed or the nodes related. Embodiments of such an identity graph can thus be searched (or navigated) to determine data associated with one or more nodes. Moreover, the similarity between, for example, the identities or entitlements may be determined using the weights of the relationships in the identity graph.

Specifically, in certain embodiments, a property graph may be thought of as a graph comprising a number of interrelated nodes. These nodes may include nodes that may have labels defining the type of the node (e.g., the type of "thing" or entity that the node represents, such as an identity, entitlement or peer group) and properties that define the attributes or data of that node. For example, the labels of the nodes of an identity graph may include "Identity", "Entitlement" or "PeerGroup". Properties of a node may include, "id", "company", "dept", "title", "location", "source" "size", "clique", "mean_similarity", or the like.

The nodes of the property graph may be interrelated using relationships that form the edges of the graph. A relationship may connect two nodes in a directional manner. These relationships may also have a label that defines the type of relationship and properties that define the attributes or data of that relationship. These properties may include an identification of the nodes related by the relationship, an identification of the directionality of the relationship or a weight or degree of affinity for the relationship between the two nodes. For example, the labels of the relationships of an identity graph may include "Similarity" or "SIM", "Has_En-

titlement" or "HAS_ENT", "Belongs_To_PeerGroup" or "BELONGS_TO_PG", or the like.

Referring then to FIG. **3**A, a graphical depiction of a portion of an example identity graph **300** is depicted. Here, nodes are represented by circles and relationships are represented by the directional arrows between the nodes. Such an identity graph **300** may represent identities, entitlements or peer groups, their association, and the degree of similarity between identities represented by the nodes. Thus, for example, the identity nodes **302***a*, **302***b* have the label "Identity" indicating they are identity nodes. Identity node **302***b* is shown as being associated with a set of properties that define the attributes or data of that identity node **302***b*, including here that the "id" of identity node **302***b* is "a123", the "company" of identity node **302***b* is "Ajax", the "dept" of identity node **302***b* is "Sales", the "title" of identity node **302***b* is "Manager", and the "location" of identity node **302***b* is "Austin, Tex.".

These identity nodes **302** of the identity graph **300** are joined by edges formed by directed relationships **312***a*, **312***b*. Directed relationship **312***a* may represent that the identity of identity node **302***a* is similar to (represented by the labeled "SIM" relationship **312***a*) the identity represented by identity node **302***b*. Similarly, directed relationship **312***b* may represent that the identity of identity node **302***b* is similar to (represented by the labeled "SIM" relationship **312***b*) the identity represented by identity node **302***a*. Here, relationship **312***b* has been assigned a similarity weight of 0.79. Notice that while these relationships **312***a*, **312***b* are depicted as individual directional relationships, such a similar relationship may be a single bidirectional relationship assigned a single similarity weight.

Entitlement nodes **304***a*, **304***b* have the label "Entitlement" indicating that they are entitlement nodes. Entitlement node **304***a* is shown as being associated with a set of properties that define the attributes or data of that entitlement node **304***a*, including here that the "id" of entitlement node **304** is "ad137", and the "source" of entitlement node **304***a* is "Active Directory". Entitlement node **304***b* is shown as being associated with a set of properties that define the attributes or data of that entitlement node **304***b*, including here that the "id" of entitlement node **304***b* is "ad179", and the "source" of entitlement node **304***a* is "Active Directory".

These entitlement nodes **304** of the identity graph **300** are joined by edges formed by directed relationships **312***c*, **312***d*. Directed relationship **312***c* may represent that the entitlement node **304***a* is similar to (represented by the labeled "SIM" relationship **312***c*) the entitlement represented by entitlement node **304***b*. Similarly, directed relationship **312***d* may represent that the entitlement of entitlement node **304***b* is similar to (represented by the labeled "SIM" relationship **312***d*) the entitlement represented by entitlement node **304***a*. Here, relationship **312***c* has been assigned a similarity weight of 0.65. Notice that while these relationships **312***c*, **312***d* are depicted as individual directional relationships, such a similar relationship may be a single bidirectional relationship assigned a single similarity weight.

Identity node **302***b* and entitlement nodes **304***a*, **304***b* of the identity graph **300** are joined by edges formed by directed relationships **316**, **316**. Directed relationships **316** may represent that the identity of identity node **302***b* has (represented by the labeled "HAS_ENT" relationships **316**) the entitlements represented by entitlement nodes **304***a*, **304***b*.

Peer group node **306***a* has the label "PeerGroup" indicating that it is a peer group node. Peer group node **306***a* is shown as being associated with a set of properties that define the attributes or data of that peer group node **306***a*, including here that the "id" of peer group node **306***a* is "pg314", the "size" of peer group node **306***a* is "287", the "clique" of peer group node **306***a* is "0.83" and the "mean_sim" or mean similarity value of peer group node **306***a* is "0.78". Identity node **302***b* and peer group node **306***a* of the identity graph **300** are joined by an edge formed by directed relationship **314***a*. Directed relationship **314***a* may represent that the identity of identity node **302***b* belongs to (represented by the labeled "BELONGS_TO_PG" relationship **314***a*) the peer group represented by peer group node **306***a*.

Peer group node **306***b* has the label "PeerGroup" indicating that it is a peer group node. Peer group node **306***b* is shown as being associated with a set of properties that define the attributes or data of that peer group node **306***b*, including here that the "id" of peer group node **306***b* is "pg763", the "size" of peer group node **306***b* is "146", the "clique" of peer group node **306***b* is "0.74" and the "mean_sim" or mean similarity value of peer group node **306***b* is "0.92". Entitlement node **304***a* and peer group node **306***b* of the identity graph **300** are joined by an edge formed by directed relationship **314***b*. Directed relationship **314***b* may represent that the identity of entitlement node **304***a* belongs to (represented by the labeled "BELONGS_TO_PG" relationship **314***b*) the peer group represented by peer group node **306***b*.

Entitlement nodes **308***a*, **308***b* have the label "Role" indicating that they are Role nodes. Role node **308***a* is shown as being associated with a set of properties that define the attributes or data of that Role node **308***a*, including here that the "id" of entitlement node **308***a* is "Role_0187". Role node **308***b* is shown as being associated with a set of properties that define the attributes or data of that role node **308***b*, including here that the "id" of role node **308***b* is "Role_3128". Directed relationship **318** may represent that the identity of identity node **302***b* has (represented by the labeled "HAS_ROLE" relationship **318**) the role represented by role node **308***a*. Directed relationship **320** may represent that the entitlement of entitlement node **304***a* is a part of or included in (represented by the labeled "PART_OF" relationship **320**) the role represented by role node **308***a*.

These role nodes **308** of the identity graph **300** are joined by edges formed by directed relationships **312***e*, **312***f*. Directed relationship **312***e* may represent that the role represented by role node **304***a* is similar to the role represented by role node **304***b*. Similarly, directed relationship **312***f* may represent that the role represented by role node **308***b* is similar to the role represented by role node **308***a*. Here, relationship **312***e* has been assigned a similarity weight of 0.34. Again, notice that while these relationships **312***e*, **312***f* are depicted as individual directional relationships, such a similar relationship may be a single bidirectional relationship assigned a single similarity weight.

FIG. **3**B is a graphical depiction of an entitlement graph and the subgraphs or clusters that may result from different pruning thresholds. In particular, entitlement graph **350** may be an initial cluster of entitlement nodes with edges having similarity weights (e.g., which may be determined as discussed) where the entitlement graph has been pruned initially and clustered according to a 0.5 pruning threshold for the similarity weight. Entitlement graph **360** is a result of pruning the entitlement graph **350** according to a higher pruning threshold of 0.8 and clustering. Here, two subgraphs **362***a*, **362***b* may result from such a pruning.

Now referring to FIGS. **3**C, **3**D and **3**E, example representations of peer groupings within identity graphs are

depicted. Here, each identity node of an identity graph is represented by a circle and each edge is represented by a line joining the nodes. In these visual depictions, the closer the nodes the higher the similarity value between the nodes. Such visual depictions when presented to a user may allow a user to better perceive the number of identities utilized by an enterprise, the relationships between those identities, the distribution of entitlements with respect to those identities or other information related to the identities or entitlements that may be utilized in identity governance and management, including for example, compliance assessment or auditing.

FIG. **4** depicts an embodiment of an interface that may be utilized by an identity management system to visually present data regarding the peer groups determined for identities within an enterprise. In this example, the enterprise has 9235 associated identities, and the interface depicts that there are 6 peer groups of those identities that have been determined based on the entitlements associated with the identities. Each of the depicted circles **410** within the interface represents one of the peer groups and displays the number of identities associated with each of those peer groups. Moreover, the size and location of each circle **410** may depict the relative size of the peer groups of the identities and the number of entitlements shared between those peer groups, or identities within those peer groups.

FIG. **5** depicts an embodiment of interface that may be utilized by an identity management system to visually present data regarding the peer groups determined for identities within an enterprise. Here, the interface may present a visual representation of the identity graph as discussed above where each identity node is represented by a circle and each edge is represented by a line joining the nodes, where the closer the nodes the higher the similarity value between the nodes. The interface may also present information regarding the number of peer groups (clusters) determined for the identity graph being presented (in this example 11).

The interface, or a portion thereof, may allow the user to navigate around the identity graph and "drill down" to obtain information on a represented node or entitlement. In the depicted example, the user has hovered above a node **510** of the identity graph and information about that identity is presented through the interface to the user. By looking at such an identity graph a user may be able to discern, for example, which identities which may be "highly contagious" or represent other identity management risks or compliance issues. An identity may be "highly contagious" or otherwise represent an identity governance risk, for example, if that identity has a number or type of entitlements such that if those identities are replicated without identity governance oversight (e.g., assigned to other users) it may cause identity governance issues such as unintended entitlement bloom.

FIG. **6** depicts an embodiment of another interface that may be utilized by an identity management system to visually present data regarding the peer groups determined for identities within an enterprise. In this example, the interface can present data regarding a particular peer group determined for an identity graph, showing, for example, the number of identities within that peer group, what the entitlements are within that peer group, what identities share those entitlements, or why those identities have been grouped together. The interface may also present a wide variety of other data regarding that peer group or identities or entitlements within that (or other) peer groups, including for example, how that peer group, identities within that peer group or other entitlements relate to each other or other determined peer groups, identities or entitlements of the enterprise. Thus, a user viewing such an interface may be able to ascertain reasons why the identities have been grouped and explore for outliers and see entitlements that these identities have in common with each other, as well as how different they are from the rest of the identities and entitlements of an enterprise. Moreover, the user may also "drill down" for more details to discover which identities included and the entitlements assigned.

FIG. **7** depicts an embodiment of still another interface that may be utilized by an identity management system to visually present data regarding the peer groups determined for identities within an enterprise. In this example, the interface can present data regarding a particular peer group (e.g., peer group **43**) determined for an identity graph, showing, for example, distributions of identities within the peer group, such as the identities of the peer group's correlation with departments, location or job title.

It will now be recalled from the discussions above, that what is desired in the context of identity governance solutions as discussed herein, are identity management solutions that allow for bottom-up role mining. While frequent pattern mining may be utilized to accomplish such role mining, such pattern mining may be combinatorial in nature and may not scale in a manner that allows for any sort of efficient implementation of role mining in real-time or other contexts where a large number of identities and entitlements may be involved.

The use of an identity graph for such role mining may, however, allow role mining that is scalable and efficient, where the role mining can be based on the nature of these identity graphs as disclosed, whereby popular or dominant entitlement patterns may be manifested as peer groups, densely connected components, cliques or pseudo cliques of identity nodes due to the dominant entitlement pattern which identity nodes within such a group may have as a result of the nature of the identity graph. Likewise, within an entitlement subgraph, a dominant or popular entitlement pattern may be manifested as a peer-group, densely connected component, clique, or pseudo-clique of entitlement nodes due to the fact that the pattern of entitlements may be shared by a sizeable set of identities as a result of the nature of the identity graph. Accordingly, the nature of an identity graph may allow the flexibility to pursue role mining on either subgraph, the identities subgraph or the entitlement subgraph (or both, or some combination, of these graphs).

Embodiments of identity management systems as disclosed may thus provide role mining based on an identity graph determined by the identity management system. In particular, embodiments as disclosed may utilize the peer grouping of an identity graph (or peer grouping of portions or subgraphs thereof) to identify roles from peer groups or the like (e.g., peer groups or other densely connected components or clusters such as cliques or pseudo-cliques).

According to embodiments, therefore, an identity graph may be constructed. A portion of the identity graph may then be determined, where this portion may include the entire identity graph, the entire entitlement subgraph, the entire identities subgraph, a portion of the entitlement subgraph or a portion of the identities subgraph. Peer groups of identities or entitlements of the portion of the identity graph can then be determined. From these peer groups a set of roles may be determined. Specifically, a set of entitlements may be determined from a peer group (e.g., of identities or entitlements), where the determined set of entitlements may represent a determined role. These roles (e.g., an identifier for the

determined role and associated entitlements) may then be stored by the identity management system for presentation to a user or other uses.

In one embodiment, the entire entitlement subgraph of the identity graph may be peer grouped substantially as discussed above, and each peer group of entitlements used as a determined role. In other words, according to these types of embodiments, the set of entitlements for each determined peer group of entitlements may represent a determined role. In certain other embodiments, these peer groups of entitlements may themselves be separated into densely connected components, cliques, or pseudo-cliques (if any exist) and the set of entitlements extracted from each of these densely connected components, cliques, or pseudo-cliques used to define the determined roles.

While the peer groups of entitlements may be utilized to perform role determination in some embodiments, it may also be observed that in many cases it is highly likely that peer groups of identities may themselves be associated with a role. This likelihood arises at because the peer grouping of identities may be based on the similarity (as represented by the edges of the identity graph) between those identities, where the similarity is, in turn, based on the number of shared entitlements. Accordingly, in certain embodiments, a set of entitlements may be extracted from each peer group of identities as determined for an identity graph (or identities subgraph), and the set of entitlements extracted from each of these identity peer groups used to define the determined roles.

To extract the set of entitlements from an identity peer group, an entitlement extraction threshold may be utilized such that an entitlement will be extracted from the identity peer group if this entitlement extraction threshold is exceeded with respect to that entitlement. This entitlement extraction threshold may be based on, for example, a threshold number, ratio or percentage of identities of the identity peer group that have that entitlement. By utilizing the identity peer groups to extract the set of entitlements for the determination of roles, in certain embodiments, the creation of the entitlement subgraph or the peer grouping of the entitlement subgraph may be avoided, substantially improving the performance of identity management systems by reducing the time, memory or processor cycles required to perform such role mining.

It will be noted here that some enterprises may have on the order of millions or more of entitlements or identities. Thus, the construction of the identity or entitlement graphs and the peer grouping of the complete set of identities or entitlements may prove prohibitive, especially in the context of real-time implementations of identity management systems and interfaces. Accordingly, other embodiments may further improve the computing performance of identity management systems by scoping the identities or entitlements to confine the graph construction, peer grouping or role determination used to determine the roles. Specifically, in one embodiment, a scoping attribute may be obtained or otherwise determined. This attribute may, for example, be provided by a user and relate to an attribute of an identity such as a title, location, department, peer group of an identity, or other data that may be obtained or associated with an identity (e.g., in association with the identity in identity management data obtained from an identity management system or in an identity graph).

Accordingly, in embodiments where an identity graph has been constructed, an identity graph may exist (e.g., have been created at a previous point) and the scoping attribute may be used to determine a subgraph of the existing identity graph to prune or to peer group in order that roles may then be determined from this subgraph. Here, a scoped identities subgraph or scoped entitlement subgraph of the identity or entitlement graphs, respectively, may be determined from the identity graph based on the scoping attribute. Specifically, in one embodiment a scoped entitlement subgraph associated with a scoping attribute may be determined from an identity graph by querying the identity graph based on the identity attribute to find the entitlement nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the identity nodes of the graph associated with the scoping attribute and determining the entitlement nodes and edges along any path of the identity graph originating with each of those scoped identity nodes. Similarly, a scoped identity subgraph associated with a scoping attribute may be determined from an identity graph by querying the identity graph based on the identity attribute to find the scoped identity nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the identity nodes of the graph associated with the scoping attribute and determining the identity nodes and edges along any path of the identity graph originating with each of those scoped identity nodes.

The pruning and peer grouping of the identities subgraph of the scoped identity subgraph or the entitlements of the scoped entitlement subgraph can then be accomplished substantially as discussed to determine peer-groups of the scoped identities or peer groups of the scoped entitlements. The peer groups of identities or entitlements can then be used for role mining. For example, a set of entitlements may be extracted from each peer group of identities as determined for the scoped identities subgraph of the identity graph, and the set of entitlements extracted from each of these identity peer groups used to define the determined roles. Similarly, the scoped entitlement subgraph of the identity graph may be peer grouped substantially as discussed above, and each peer group of entitlements used as a determined role.

In certain embodiments, such as when no identity graph has been previously constructed, the scoping attribute may be used in the construction of an identity graph. In these types of embodiments, the identity management data of obtained from the identity management (e.g., a most recent snapshot of the identity management data) may be searched based on the scoping attribute to determine the identities of the identity management data associated with that scoping attribute (e.g., identities having that title, location, department, etc.) and the entitlements associated with those determined identities. Once this scoped set of identities and entitlements is determined, an identity graph may be generated from the scoped set of identities and entitlements substantially as discussed above, where a node of the graph is created for each scoped identity and entitlement, and weighted edges are constructed between every pair of identity nodes that shares at least one entitlement and between every pair of entitlement nodes that shares at least one identity. The pruning and peer grouping of the identities subgraph or the entitlements subgraph of the scoped identity graph can then be accomplished substantially as discussed to determine peer-groups of the scoped identities or peer groups of the scoped entitlements. The peer groups of identities or entitlements can then be used for role mining.

For example, a set of entitlements may be extracted from each peer group of identities as determined for the identities subgraph of the scoped identity graph, and the set of entitlements extracted from each of these identity peer

groups used to define the determined roles. Similarly, the entitlement subgraph of the scoped identity graph may be peer grouped substantially as discussed above, and each peer group of entitlements used as a determined role. It will be noticed here, that in instances where it is desired to only utilize a entitlement subgraph or an identity subgraph for role mining, only an identities subgraph or an entitlements subgraph may be created from the scoped set of identities or entitlements, and this subgraph pruned and clustered as described to yield the desired identity peer groups or entitlement peer groups to utilize for such role mining.

Referring to FIG. **8**, a distributed networked computer environment including an identity management system with one embodiment of a role miner is depicted. As discussed above, the networked computer environment may include an enterprise computing environment **800** including a number of computing devices or applications that may be coupled over a computer network **802** or combination of computer networks. Enterprise environment **800** may thus include a number of resources, various resource groups and users associated with an enterprise. Users may have various roles, job functions, responsibilities, etc. to perform within various processes or tasks associated with enterprise environment **800**.

Users may access resources of the enterprise environment **800** to perform functions associated with their jobs, obtain information about enterprise **800** and its products, services, and resources, enter or manipulate information regarding the same, monitor activity in enterprise **800**, order supplies and services for enterprise **800**, manage inventory, generate financial analyses and reports, or generally to perform any task, activity or process related to the enterprise **800**. Thus, to accomplish their responsibilities, users may have entitlements to access resources of the enterprise environment **800**. These entitlements may give rise to risk of negligent or malicious use of resources.

Specifically, to accomplish different functions, different users may have differing access entitlements to differing resources. Some access entitlements may allow particular users to obtain, enter, manipulate, etc. information in resources which may be relatively innocuous. Some access entitlements may allow particular users to manipulate information in resources of the enterprise **800** which might be relatively sensitive. Some sensitive information can include human resource files, financial records, marketing plans, intellectual property files, etc. Access to sensitive information can allow negligent or malicious activities to harm the enterprise itself. Access risks can thus result from a user having entitlements with which the user can access resources that the particular user should not have access to for other reasons. Access risks can also arise from roles in enterprise environment **800** which may shift, change, evolve, etc. leaving entitlements non optimally distributed among various users.

To assist in managing the entitlements assigned to various users and more generally in managing and assessing access risks in enterprise environment **800**, an identity management system **850** may be employed. Such an identity management system **850** may allow an administrative or other type of user to define one or more identities and one or more entitlements and associate these identities with entitlements using, for example, an administrator interface **852**. Moreover, an identity management system **850** may allow such a user to define one or more roles for the enterprise, where these defined enterprise roles are defined as collections of access entitlements or access profiles and may be assigned to identities through the identity management system **850** based on specific rules of the enterprise in terms of the identity's attributes, their expected responsibilities within the organization, or other criteria. These enterprise roles as defined by the user through the identity management system **850** may thus define an ideal or 'golden' state of the roles of an enterprise.

Examples of such identity management systems are Sailpoint's IdentityIQ and IdentityNow products. Note here, that while the identity management system **850** has been depicted in the diagram as separate and distinct from the enterprise environment **800** and coupled to enterprise environment **800** over a computer network **804** (which may the same as, or different than, network **802**), it will be realized that such an identity management system **850** may be deployed as part of the enterprise environment **800**, remotely from the enterprise environment, as a cloud based application or set of services, or in another configuration.

The identity management system **850** may thus store identity management data **854**. The identity management data **854** stored may include a set of entries, each entry corresponding to and including an identity (e.g., alphanumerical identifiers for identities) as defined and managed by the identity management system, a list or vector of entitlements (e.g., alphanumeric identifiers for entitlements) assigned to that identity by the identity management system, a list or vector of enterprise roles assigned to that identity, and a timestamp at which the identity management data was collected from the identity management system. Other data could also be associated with each identity, including data that may be provided from other systems such as a title, location or department associated with the identity. The set of entries may also include entries corresponding to entitlements and roles, where each entry for a role may include the role identifier (e.g., alphanumerical identifier or name for the role) and a list or vector of the entitlements associated with each role. Other data could also be associated with each role, such as a title, location or department associated with the role.

Accordingly, the collectors **856** of the identity management system **850** may obtain or collect event data from various systems within the enterprise environment **800** and process the event data to associate the event data with the identities defined in the identity management data **854** to evaluate or analyze these events or other data in an identity management context. As part of a robust identity management system, it is desirable to analyze the identity management data **854** associated with an enterprise **800**. Accordingly, an identity management system **860** may include a harvester **862** and a graph generator **864**. The harvester **862** may obtain identity management data **854** from one or more identity management systems **850** associated with enterprise **800**. Graph generator **864** may allow an identity graph or subgraphs thereof to be generated from the obtained identity management data **854** and stored in graph data store **866**. Interfaces **868** of the identity management system **860** or interface **858** may use a graph in the graph data store **866** or associated peer groups to present one or more interfaces which may be used for risk assessment, including the presentation of roles mined from such graphs.

Additionally, a user may interact with the identity management system **850** through a user interface **858** to access or manipulate data on identities, entitlements, events, roles or generally perform identity management with respect to enterprise environment **800**. As but one example, as the roles, entitlements and identities of an enterprise evolve they may stray in substantial and detrimental ways from an ideal state, or other identity governance desires, of the enterprise.

Users of an identity management system may thus wish to determine a current data-driven assessment of the current role structure for their enterprise.

By determining a current snapshot of the roles mined from an actual state of their identity governance structure, the 'golden' enterprise roles as defined by the users of the enterprise may be compared with the mined roles to reduce discrepancies therebetween, including for example, the identification of new roles, the evolution of the enterprise defined roles to match the evaluation of the actual role structure (e.g., the mined roles), or the performance housekeeping on the assignment of entitlements or roles within the enterprise to more particularly tailor the actual role structure to the ideal role structure. Additionally, by viewing the mined roles extraneous, singleton or outlier entitlements that have been deprecated or are in need of certification may be identified.

Accordingly, it is desirable for identity management solutions to offer a role mining capability whereby collections of entitlements may be ascertained from the identity management data associated with enterprise. Embodiments of identity management system **850** may thus provide a role mining tool through the user interface **858**. In this manner, a user may be presented with the ability to perform role mining through the user interface **858** (or interface **868**), along with an optional attribute or criteria to scope the mining of the roles. One or more interfaces with the results of the role mining can then be determined and presented to the user through the role mining tool of user interface **858**.

To provide such a role mining tool, identity management system **860** may include role miner **880**. Role miner **880** may include an interface **882**. When a request for role mining is received from the user through the user interface **858**, a request to perform role mining can be submitted to the role miner **880** from the identity management system **850** (e.g., or user interface **858** or other component of identity management system **850**) through the interface **882**, where the request may include zero or more scoping attributes that may have been provided by the user through the user interface. The request may include other criteria or attributes, such as a pruning threshold to utilize when creating an identity graph or subgraph or a role size which may be used to determine which graphs or subgraphs to utilize when performing role extraction. Note here, that while the identity management system **850** has been depicted in the diagram as separate and distinct from the identity management system **860** and coupled to identity management system **860** over a computer network **804**, it will be realized that such an identity management system **850** and identity management system **860** may be deployed as part of the same identity management system or different identity management system, as a cloud based application or set of services, or in another configuration entirely.

As such, when a request for role mining and the zero or more associate scoping (or other) attributes are received through the role minter interface **882**, the role miner **880** may preform role mining as discussed. In particular, in one embodiment, the role miner may determine if there is an identity graph in graph data store **866**, or if the identity graph currently in graph data store **866** was created within some previous time window (e.g., last hour, last twenty four hours, last week, etc.). If there is an identity graph available (e.g., if one exists in the graph data store **866** or was created within the time window), the role miner **880** can determine if a scoring attribute was received with the role mining request. If no scoring attribute was received, the available identity graph may be used for role mining. If, however, a scoring attribute was received and an identity graph is available, the existing identity graph can be scoped based on the received scoring attribute and the type of role mining to be performed. As discussed, the role miner **880** may perform role mining based on an identities subgraph, an entitlement subgraph, or some combination according to various embodiments. The type of role mining to be performed may, for example, be configured by an administrator of the identity management system **860** or may be specified by a user in a request for role mining using the role mining interface **858**.

Accordingly, in embodiments where an identity graph has been constructed, an identity graph may exist (e.g., have been created at a previous point) in the graph data store **866** and the scoping attribute may be used to determine a subgraph of the existing identity graph to use as a role mining graph. Peer groups may be determined from that role mining graph in order that roles may be then be determined from this subgraph. Here, a scoped identities subgraph or scoped entitlement subgraph of the identity graph may be determined from the identity graph based on the scoping attribute and the type of role mining to be performed.

Specifically, in one embodiment a scoped entitlement subgraph associated with a scoping attribute may be determined from an identity graph by querying the identity graph in the graph data store **866** based on the scoping attribute to find the entitlement nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the identity nodes of the graph associated with the scoping attribute and determining the entitlement nodes and edges along any path of the identity graph originating with each of those scoped identity nodes.

Similarly, a scoped identity subgraph associated with a scoping attribute may be determined from an identity graph in the graph data store **866** by querying the identity graph based on the identity attribute to find the scoped identity nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the identity nodes of the graph associated with the scoping attribute and determining the identity nodes and edges along any path of the identity graph originating with each of those scoped identity nodes.

Once a graph to utilize for role mining is determined (e.g., the entire available identity graph, the scoped identities subgraph, the scoped entitlement subgraph or some combination), the peer groups of this role mining graph may be determined. In some cases, in instances where an identity graph is available the identity graph may also have been previously pruned and peer grouped. Thus, the peer groups associated with the role mining graph to use for role mining may be determined by accessing the peer groups previously determined for the identity graph that are associated with the nodes of the role mining graph. For example, if identity peer groups are to be utilized for role mining, the identity peer groups associated with the identity nodes of the role mining graph may be determined from the previously determined entitlement peer groups for the identity graph. Likewise, if entitlement peer groups are to be utilized for role mining, the entitlement peer groups associated with the entitlement nodes of the role mining graph may be determined from the previously determined entitlement peer groups for the identity graph.

If however, peer groups have not been determined for the available identity graph, the role mining graph (e.g., the entire available identity graph, the scoped identities subgraph, the scoped entitlement subgraph or some combina-

tion) may be peer grouped as discussed above based on the type of role mining to be performed. For example, the pruning and peer grouping of a role mining graph comprising a scoped identities subgraph or a scoped entitlement subgraph can be accomplished substantially as discussed to determine identity peer-groups of the scoped identities or entitlement peer groups of the scoped entitlements.

These peer groups of identities or entitlements can then be used by the role miner for role mining. For example, a set of entitlements may be extracted from each peer group of identities determined for a role mining graph (e.g., a scoped identities subgraph of the identity graph), and the set of entitlements extracted from each of these identity peer groups used to define the determined roles. To extract the set of entitlements from an identity peer group, an entitlement extraction threshold may be utilized such that an entitlement will be extracted from the identity peer group if this entitlement extraction threshold is exceeded with respect to that entitlement. This entitlement extraction threshold may be based on, for example, a threshold number, ratio or percentage of identities of the identity peer group that have that entitlement. Similarly, in cases where the role mining graph is a scoped entitlement subgraph of the identity graph, the entitlements of the role mining graph may be peer grouped substantially as discussed above, and each peer group of entitlements used as a determined role.

In instances where the role miner **880** receives a request for role mining and zero or more associated scoping attributes and there is no identity graph currently in graph data store **866** (or the graph in the graph data store **866** was created outside of some previous time window), a role mining graph may be determined from the identity management data **854** and the type of role mining to be performed.

In particular, according to certain embodiments the role miner **880** can determine if a scoping attribute was received with the received request. If there is no scoping attribute, identity management data **854** may be obtained and an identity graph (or identities subgraph or entitlements subgraph) constructed as previously discussed. The type of graph constructed for use as a role mining graph may be dependent on the type of role mining to be performed as discussed.

However, if a scoping attribute was received, the identity management data **854** obtained from the identity management system **850** (e.g., a most recent snapshot of the identity management data) may be scoped based on the attribute by searching the identity management data **854** based on the scoping attribute to determine the identities of the identity management data **854** associated with that scoping attribute (e.g., identities having that title, location, department, etc.) and the entitlements associated with those determined identities determined.

Once this scoped set of identities and entitlements is determined, a role mining graph (e.g., an identity graph, identities subgraph or entitlements subgraph) may be generated from the scoped set of identities and entitlements substantially as discussed above, where a node of the graph is created for each scoped identity and entitlement, and weighted edges are constructed between every pair of identity nodes that shares at least one entitlement and between every pair of entitlement nodes that shares at least one identity.

The pruning and peer grouping of the role mining graph constructed from the scoped identity management data (e.g., the identity graph, entitlement subgraph, identities subgraph) can then be accomplished substantially as discussed to determine peer-groups of the scoped identities or peer groups of the scoped entitlements. The determined peer groups of identities or entitlements can then be used for role mining depending on the type of role mining to be accomplished. If a pruning threshold has been provided by a user this pruning threshold may be utilized in pruning the role mining graph. Additionally, if a role size has been provided this may be utilized during role mining such that role will only be extracted from subgraphs of the role mining graph that have a number of nodes exceeding the role size.

Once the role miner **880** has determined the set of entitlements comprising each of the mined roles, these mined roles may be presented to the user through interface **858** of the identity management system **850**. These roles may, for example, be assigned an identifier by the role miner **880** (e.g., an assigned alphanumerical identifier or a semantic identifier that may be determined, for example, from identities or entitlements associated with the role). The roles may then be presented graphically in an interface **858** with which the user may interact to determine additional or different data about the presented roles.

FIG. **9** depicts one embodiment of a method for role mining that may be used in embodiments of an identity management system such as those disclosed herein.

Initially, at step **910** a request to perform role mining can be received, where the request may include zero or more scoping attributes (e.g., that may have been provided by the user or otherwise determined). At step **920** it can be determined if an identity graph is available. This determination may involve determining if an identity graph has been previously created or is a previously create identity graph was created within some previous time window (e.g., last hour, last twenty four hours, last week, etc.).

If there is an identity graph available (Yes branch of step **920**), it can then be determined at step **930** determine if a scoping attribute was received with the role mining request. If no scoring attribute was received (No branch of step **930**), the available identity graph may be used for role mining. If, however, a scoring attribute was received and an identity graph is available (Yes branch of step **930**), the existing identity graph can be scoped at step **940** based on the received scoring attribute and the type of role mining to be performed. As discussed, the role mining may be performed based on an identities subgraph, an entitlement subgraph, or some combination according to various embodiments.

Accordingly, in embodiments where an identity graph has been constructed, an identity graph may exist (e.g., have been created at a previous point) and the scoping attribute may be used to determine a subgraph of the existing identity graph to use as a role mining graph. Peer groups may be determined from that role mining graph in order that roles may be then be determined from this subgraph. Here, at step **940**, a scoped identities subgraph or scoped entitlement subgraph of the identity graph may be determined from the identity graph based on the scoping attribute and the type of role mining to be performed.

Specifically, in one embodiment a scoped entitlement subgraph associated with a scoping attribute may be determined from an identity graph by querying a previously created identity graph based on the scoping attribute to find the entitlement nodes and edges associated with the scoping attribute. Similarly, a scoped identity subgraph associated with a scoping attribute may be determined from a previously created identity graph by querying the identity graph based on the identity attribute to find the scoped identity nodes and edges associated with the scoping attribute.

Once a graph to utilize for role mining at is determined at step **940**, the peer groups of this role mining graph may be

determined at step **950**. In many cases, in instances where an identity graph is available the identity graph may also have been previously pruned and peer grouped. Thus, the peer groups associated with the role mining graph to use for role mining may be determined by accessing the peer groups previously determined for the identity graph that are associated with the nodes of the role mining graph. If however, peer groups have not been determined for the available identity graph, the role mining graph (e.g., the entire available identity graph, the scoped identities subgraph, the scoped entitlement subgraph or some combination) may be peer grouped as discussed above based on the type of role mining to be performed. For example, the pruning and peer grouping of a role mining graph comprising a scoped identities subgraph or a scoped entitlement subgraph can be accomplished substantially as discussed to determine identity peer-groups of the scoped identities or entitlement peer groups of the scoped entitlements.

These peer groups of identities or entitlements can then be used by for role mining at step **960** by extracting the roles from these peer groups. For example, a set of entitlements may be extracted from each peer group of identities determined for a role mining graph and the set of entitlements extracted from each of these identity peer groups used to define the determined roles. Similarly, in cases where the role mining graph is a scoped entitlement subgraph of the identity graph, the entitlements of the role mining graph may be peer grouped substantially as discussed above, and each peer group of entitlements used as a determined role.

Returning to step **920**, in instances where a request for role mining and zero or more associated scoping attributes is received and there is no available identity graph (No branch of step **920**) a role mining graph may be determined from the identity management data and the type of role mining to be performed.

In particular, at step **970** it can be determined if a scoping attribute was received with the received request. If there is no scoping attribute (NO branch of step **970**), identity management data may be obtained at step **980** and a role mining graph (e.g., identity graph, identities subgraph or entitlements subgraph) constructed as previously discussed at step **982**. The type of graph constructed for use as a role mining graph may be dependent on the type of role mining to be performed.

However, if a scoping attribute was received (Yes branch of step **970**), scoped identity management data may be obtained at step **984** based on the attribute by searching the identity management data based on the scoping attribute to determine the identities of the identity management data associated with that scoping attribute (e.g., identities having that title, location, department, etc.) and the entitlements associated with those determined identities determined.

Once this scoped set of identities and entitlements is determined, a role mining graph (e.g., an identity graph, identities subgraph or entitlements subgraph) may be generated from the scoped set of identities and entitlements at step **982**, whereby a node of the graph is created for each scoped identity and entitlement, and weighted edges are constructed between every pair of identity nodes that shares at least one entitlement and between every pair of entitlement nodes that shares at least one identity.

The pruning and peer grouping of the role mining graph constructed from the obtained identity management data (e.g., the identity graph, entitlement subgraph, identities subgraph) can then be accomplished at step **986** substantially as discussed to determine peer-groups of the identities or peer groups of the entitlements of the determined role mining graph. The determined peer groups of identities or entitlements can then be used for role mining depending on the type of role mining to be accomplished at step **960**.

Once the roles have been determined (e.g., the set of entitlements comprising each of the mined roles), these mined roles may be presented to the user through interface of the identity management system at step **990**. These roles may, for example, be assigned an identifier and presented graphically in an interface with which the user may interact to determine additional or different data about the presented roles.

FIGS. **10-12** depict embodiments of these types of interfaces that may be utilized by embodiments of an identity management system as disclosed herein. Looking first at FIG. **10**, one embodiment of an interface for an identity management system that may be utilized in association with role mining is depicted. Here, the interface **1010** may include an area **1020** that allows specification of a scoping attribute or other criteria associated with role mining. In the depicted example, the area **1020** may allow a user selection of an identity peer group, a pruning threshold to be used and a minimum role size. A graph display area **1030** may display the graphs or subgraphs of identity nodes and similarity relationships resulting from application of the attributes selected by the user in area **1020**.

Specifically, in one embodiment, when a user selects a particular peer group in area **1020** the identity graph or identity management data maintained by the identity management system may be scoped based on the peer group selected by the user. An identities graph may be created by the identity management system using the identities of the selected peer group and edges between the identity nodes based on shared entitlements as described. The edges of this identities graph can then be pruned according to the user selected pruning threshold in area **1020** and the pruned graph displayed in graph display area **1030**.

Area **1040** can display a view of roles (here referred to as components) mined from the graph created and displayed in graph display area **1030**. In one embodiment, once the identities graph is created, the identity management system may perform role mining on the identities graph as discussed above. In particular, according to one embodiment, the identity management system will determine each distinct subgraph of the determined identities graph, and for those subgraphs, determine which, if any, of those subgraphs includes a number of nodes greater than the minimum role size defined by the user in area **1020**. The roles can then be mined from any of these determined subgraphs.

Area **1040** will then display an indicator for each of the subgraphs of the graph displayed in graph **1030** from which a role was mined. In one embodiment, the role may be assigned an identifier and an icon (e.g., circle) representing the role may be displayed along with the identifier for the role in area **1040**. The size of the icon may, for example, be reflective of the number of nodes or size of the subgraph from which the associated role was mined.

In this example, a user has selected a peer group "wcP3054" and defined a pruning threshold of "0.6" and a minimum role size of 1 in area **1020**. Thus, an identities graph may be generated by the identity management system using the identities associated with the peer group "wcP3054" using a pruning threshold of 0.6 for the edges. This identities graph is displayed in graph display area **1030**. Moreover, here there is only subgraph of the identities graph and it has a greater number of identity nodes than the minimum role size of 1 specified by the user. Thus, the identities management system may perform role mining on

this identities graph, assign the identifier "C-18" to the mined role and display an icon **1042** in area labeled with the role identifier ("C-18") with a size reflective of the number of nodes of the identity graph represented in graph display area **1030** from which the role was mined.

Moving to FIG. **11**, here, the user has selected the same peer group "wcP3054" and defined a pruning threshold of "0.8" and a minimum role size of 1 in area **1020**. Thus, an identities graph may be generated by the identity management system using the identities associated with the peer group "wcP3054" using a pruning threshold of 0.8 for the edges. This identities graph is displayed in graph display area **1030**. Here, however, as the edges have been pruned according to a higher pruning threshold (e.g., 0.8) there are 6 subgraphs **1104** of the identities graph that have a greater number of identity nodes than the minimum role size of 1 specified by the user. Thus, the identities management system may perform role mining on each of these subgraphs **1104**, assign identities to each of the mined roles and display an associated icon **1142** with the assigned label in area **1040**, where the icon **1142** may have a size reflective of the number of nodes in the associated subgraph. Here, for example, icon **1142***a* may be associated with subgraph **1104***a*, icon **1142***b* associated with subgraph **1104***b*, icon **1142***c* associated with subgraph **1104***c*, etc.

Continuing with the same example, in FIG. **12** the user has selected the same peer group "wcP3054" and defined a pruning threshold of "0.8.' However, here the user has defined a minimum role size of 10 in area **1020**. Thus, an identities graph may be generated by the identity management system using the identities associated with the peer group "wcP3054" using a pruning threshold of 0.8 for the edges. This identities graph is displayed in graph display area **1030**. Here, as in FIG. **11**, there are 6 subgraphs **1204**. However, as the user has defined a minimum role size of 10 in the example depicted in FIG. **12**, the identities management system may only perform role mining on each of these subgraphs **1204***a*, **1204***b* that have more than 10 identity nodes. The identity management system can then assign identities to each of the mined roles and display an associated icon **1242** with the assigned label in area **1040**, where the icon **1242** may have a size reflective of the number of nodes in the associated subgraph. Here, for example, icon **1242***b* may be associated with subgraph **1204***a* and icon **1242***b* associated with subgraph **1204***b*.

FIG. **13** depicts an embodiment of an interface that may be utilized by an identity management system to display a distribution of entitlements within a particular role. Specifically, in the depicted embodiment, when a user selects a particular role (e.g., within an interface presented by the identity management system), the user may be presented with an interface such as that in FIG. **13** whereby a list of entitlements of the role and the distribution of those entitlements may be presented to a user in both a textual manner and through a visual depiction, such as a histogram or the like.

FIG. **14** depicts another embodiment of an interface that may be utilized by an identity management system to display data regarding determined roles. Here, the interface may be a Sankey chart showing which roles (e.g., Role 6, Role 5 and Role 34) include certain attributes (e.g., Engineering, Sales, Software Engineer, Data Scientist and Product Manager).

As can be seen then, according to embodiments of an identity management system, an identity graph may include nodes representing roles, where those roles may be defined based on identity management data obtained from an enterprise, roles defined by a user associated with an enterprise (e.g., using a role definition interface) or determined from role mining, or from another source altogether. It is thus desirable for identity management systems to offer role assessment capabilities whereby roles comprising collections of entitlements may be ascertained from the identity management data associated with enterprise and an assessment metric (also refer to as a score) for a set of these roles may be determined, where the metric is a reflection, for example, of the quality or health (used herein interchangeably) of the structure of the set of roles.

Accordingly, to ameliorate or address these issues, among other ends, embodiments of the identity management systems disclosed herein may utilize a network graph approach to improve identity governance, including the assessment of roles associated with the identity management data of an enterprise. Specifically, embodiments of identity management systems as disclosed may provide role assessment based on a network graph that includes roles of an enterprise. Embodiments may thus generate a network identity (property) graph that includes nodes for identities, entitlements, roles or other identity management artifacts of an enterprise. Such a network identity graph may be, or may include, a role graph having nodes representing roles associated with the enterprise and edges representing similarities between the roles (e.g., represented by the nodes). These edges may comprise a similarity weight determined, based on, for example, shared entitlements between the roles or by concurrent identities (e.g., a number of identities that share those roles).

Specifically, in many instances, in the context of an enterprise there may be what are referred to as multi-dimensional roles. A multi-dimensional role may be instances of similar roles that may vary slightly according to some criteria. For example, if an enterprise has many different locations, a role in one location may be very similar to a role in another location. Thus, administrators or others concerned with identify governance within an enterprise, or compliance of an enterprise with identity management goals or requirements, may desire to validate or otherwise assess the role structure of an enterprise (or portions thereof) to determine the quality or health of these roles. By assessing the health of the roles structure, such metrics may be useful for compliance purposes or to assist in optimizing the role structure or more generally streamlining role management for the enterprise.

Moreover, by identifying roles that may be strongly similar or otherwise closely aligned, efficiencies with respect to management of these roles may be achieved. For example, in some cases, roles that have similar sets of entitlements may be consolidated (e.g., merged) or some of the roles eliminated. As another alternative, roles that share a similar group of identities (e.g., where the same set of identities share a set of roles) may be bundled together and an overarching role (referred to as a portfolio role) may be defined such that the bundle of similar roles may be managed as a group using the portfolio role. Thus, using embodiments, the actual scope of identities (e.g., a user population) for which roles can be consolidated to reduce use of resources in role management for that specific population and defining or assigning roles for that population. More generally, then, by reducing the number of roles or the interactions with these roles, the number of both computing resources and man hours required for such identity governance may be reduced, along with the commensurate cost to the enterprise of such identity management.

In one embodiment, for example, a role graph may be an access role graph that is a role graph modeled in terms of entitlement (e.g., access) similarities between all the roles. The edges of the access role graph represent an access similarity relationship between two roles (e.g., nodes representing the roles) joined by the edge of the graph. A weight may be computed for the access similarity relationship based on the entitlements shared between the two roles and the number of entitlements the roles include. Roles with similar entitlements or access patterns may thus cluster close together on the access graph. Embodiments of these access role graphs may give high-level of abstractions on the overall access model of an enterprise while accurately reflecting the global role (access) structure. As such, these access role graphs may be useful, for example, as a "role provisioning Quality Assessment" tool indicating overall well-being of an enterprises role structure, in recommending consolidation of redundant roles, or verifying how new roles may fit in the current access model.

As another embodiment, for example, a role graph may be an concurrency graph that is modeled in terms of concurrent identities shared between roles. The edges of the concurrency graph represent an concurrency similarity relationship between two roles (e.g., nodes representing the roles) joined by the edge of the graph. A weight may be computed for the concurrency similarity relationship based on the number of identities which share those roles and the number of identities that have those roles. Roles with high concurrency with one another cluster closer together on the concurrency graph. Moreover, the concurrency graph may be filtered based on the number of supporting identities (e.g., the number of identities that include both roles). This support (also referred to as the concurrent or concurrency count) thus determines the significance of the computed concurrency weights, by allowing the concurrency graph to filter out highly concurrent roles that share only few identities, thus rendering more meaningful representation of the concurrency graph. As such, these concurrency graphs may be useful as a "role-profiling assistant" identifying concurrent patterns of peer access, simplifying business rules, or surfacing potential profiles for new joiners. These concurrency graphs may also allow users to dive deeper and profile roles within units of an enterprise when applied with scoping of the concurrency graph.

Moreover, according to embodiments, various metrics may be determined for assessing the quality or health of the role structure of an enterprise based on an access role graph or a concurrency role graph. Specifically, optimal (e.g., ideal) network or graph topologies for access and concurrency graphs can be inferred. Graph based metrics may thus provide a starting point to standardize quality scoring for role structures and access models. In one embodiment, a combination of graph based metrics may be utilized to measure a role graph structure with respect to an ideal graph topology optimized for the enterprise. Such a scoring system allows personalization taking into account the trade-off between compliance-driven and enablement-driven governance strategies. Thus role data, including for example, visual depictions of role graphs for the enterprise or quality assessment scores may be presented to a user through embodiments of the identity management systems as depicted herein.

Turning to FIG. **15** then, a distributed networked computer environment including an identity management system with one embodiment of a role assessor is depicted. As discussed above, the networked computer environment may include an enterprise computing environment **1500** including a number of computing devices or applications that may be coupled over a computer network **1502** or combination of computer networks. Enterprise environment **1500** may thus include a number of resources, various resource groups and users associated with an enterprise. Users may have various roles, job functions, responsibilities, etc. to perform within various processes or tasks associated with enterprise environment **1500**.

To assist in managing the entitlements assigned to various users and more generally in managing and assessing access risks in enterprise environment **1500**, an identity management system **1550** may be employed. Such an identity management system **1550** may allow an administrative or other type of user to define one or more identities and one or more entitlements and associate these identities with entitlements using, for example, an administrator interface **1552**. Moreover, an identity management system **1550** may allow such a user to define one or more roles for the enterprise, where these defined roles are defined as collections of access entitlements or access profiles and may be assigned to identities through the identity management system **1550** based on specific rules of the enterprise in terms of the identity's attributes, their expected responsibilities within the organization, or other criteria. Identity management system **1550** or **1560** may also allow roles to be mined and defined in this manner. Identity management system **1550** may, in many respect, function similarly to other embodiments of identity management systems disclosed herein and such similar functionality will not be described further for the sake of conciseness.

The identity management system **1550** may thus store identity management data **1554**. The identity management data **1554** stored may include a set of entries, each entry corresponding to and including an identity (e.g., alphanumerical identifiers for identities) as defined and managed by the identity management system, a list or vector of entitlements (e.g., alphanumeric identifiers for entitlements) assigned to that identity by the identity management system, a list or vector of enterprise roles assigned to that identity, and a timestamp at which the identity management data was collected from the identity management system. The set of entries may also include entries corresponding to entitlements and roles, where each entry for a role may include the role identifier (e.g., alphanumerical identifier or name for the role) and a list or vector of the entitlements associated with each role. Other data could also be associated with each role, such as a title, location or department associated with the role.

Accordingly, graph generator **1564** may obtain identity management data **1554** from one or more identity management systems **1550** associated with enterprise **1500**. Graph generator **1564** may allow an identity graph or subgraphs thereof to be generated from the obtained identity management data **1554** and stored in graph data store **1566**. In one embodiment, as part of a generated identity graph, or as separate graphs, graph generator may generate one or more role graphs.

Again, these role graphs may be subgraphs of an identity graph, or may be separately generated and stored, by the graph generator **1564**. In one embodiment, for example, graph generator **1564** may generate an access role graph that is a role graph modeled in terms of entitlement (e.g., access) similarities between all the roles. The edges of the access role graph represent an access similarity relationship between two roles (e.g., nodes representing the roles) joined by the edge of the access role graph. Thus, in one embodiment, an access role graph may be generated from identity

management data obtained from the enterprise. This access role graph may be, for example, be generated as part of an identity graph and may be generated in association with such an identity graph by graph generator **1564** (and may thus be a subgraph of such an identity graph).

Specifically, in generating such an identity graph, each of the roles from the most recently obtained identity management data may be determined and a node of the graph created for each role. An edge is constructed between each role node (node representing a role) and each entitlement node representing an entitlement included in that node, where that edge may represent a relationship that indicates the role includes that entitlement. An edge of the identity graph may also be constructed between each identity node (node representing an identity) and each role node representing a role that has been assigned to that identity, where that edge may represent a relationship that indicates the identity has that role.

There may also be an edge constructed between role nodes that represents a relationship (referred to as an access similarity) between the roles represented by the nodes based on the number of entitlements shared by the roles represented by those nodes (e.g., where each of the role nodes has an edge in the graph to the same entitlement node representing that each role includes the entitlement represented by the entitlement node). A weight may be computed for the access similarity relationship based on the entitlements shared between the two roles and the number of entitlements each of the roles include.

Such a weight for an access similarity relationship may be generated to represent a degree of similarity between the roles of the respective nodes joined by that edge based on the number of shared entitlements. In one embodiment, for example, using a proper similarity function (e.g., Jaccard similarity). In one embodiment, the Jaccard similarity for an access relationship between two role nodes may be determined by the Intersection(entitlements of the two roles represented by the role nodes)/Union(entitlements of the two roles represented by the role nodes). In this manner then, a generated identity graph may include an access role graph that is a role graph modeled in terms of entitlement (e.g., access) similarities between the roles.

Similarly, there may be an edge constructed between role nodes that represents a relationship (referred to as a concurrent similarity) between the roles represented by the nodes based on the number of identities that shared by the roles (e.g., concurrent identities) represented by those nodes (e.g., where each of the role nodes has an edge in the graph to the same identity node representing that the identity includes that role). A weight may be computed for the concurrent similarity relationship based on the identities shared between the two roles and the number of identities having each of the roles.

Such a weight for a concurrent similarity relationship may be generated to represent a degree of similarity between the roles of the respective nodes joined by that edge based on the number of shared identities. For example, a weight for a concurrent similarity relationship may be generated using a proper similarity function (e.g., Jaccard similarity). In one embodiment, the Jaccard similarity for a concurrent similarity relationship between two role nodes may be determined as the Intersection(identities having both roles)/Union(identities that have either of the roles). In this manner then, a generated identity graph may include a concurrency role graph that is a role graph that modeled in terms of concurrent identities shared between roles. It will thus be noted that a generated identity graph may include a role graph (e.g., as a subgraph of the identity graph), where that role graph may include one of, or both of, an access role graph and a concurrency role graph.

As noted, a user may interact with the identity management system **1550** through a user interface **1558** to access or manipulate data on identities, entitlements, events, roles or generally perform identity management with respect to enterprise environment **1500**. As but one example, as the roles, entitlements and identities of an enterprise evolve they may stray in substantial and detrimental ways from an ideal state, or other identity governance desires, of the enterprise. Users of an identity management system may thus wish to determine a current data-driven assessment of the current role structure for their enterprise.

Specifically, in many instances, in the context of an enterprise **1500** there may be what are referred to as multi-dimensional roles. A multi-dimensional role may be instances of similar roles that may vary slightly according to some criteria. For example, if an enterprise has many different locations, a role in one location (e.g., a software developer role in Austin, Tex.) may be very similar to a role in another location (a role for a software developer in San Jose, Calif.). In other words, a software developer in either location may require access to a substantially similar set of entitlements, however, since the creators of such roles (which may be, for example, in those two different locations) may have no visibility or access into the roles structure of the enterprise generally, two (or more) different roles may be created, despite the fact that these roles may be substantially similar (e.g., comprise similar entitlements) or, in certain cases, may even be the same. Thus, administrators or others concerned with identify governance within an enterprise, or compliance of an enterprise with identity management goals or requirements, may desire to validate or otherwise assess the role structure of an enterprise (or portions thereof) to ascertain, or determine the quality or health of the roles of an enterprise.

Likewise, by identifying roles that may be strongly similar or otherwise closely aligned, efficiencies with respect to management of these roles may be achieved. For example, in some cases, roles that have similar sets of entitlements may be consolidated (e.g., merged) or some of the roles eliminated. As another alternative, roles that share a similar group of identities (e.g., where the same set of identities share a set of roles) may be bundled together and an overarching role (referred to as a portfolio role) may be defined such that the bundle of similar roles may be manage as a group using the portfolio role. Thus, using embodiments, the actual scope of identities (e.g., a user population) for which roles can be consolidated to reduce use of resources in role management for that specific population and defining or assigning roles for that population. More generally, then, by reducing the number of roles or the interactions with these roles, the number of both computing resources and man hours required for such identity governance may be reduced, along with the commensurate cost to the enterprise of such identity management.

Accordingly, is desirable for identity management solutions to offer role assessment capabilities whereby roles comprising collections of entitlements may be ascertained from the identity management data associated with enterprise **1500**. Embodiments of identity management system **1550** may thus provide a role validation tool through the user interface **1558** or interface **1568**. In this manner, a user may be presented with the ability to perform role validation through the user interface **1558** (or interface **1568**), along with an optional attribute or criteria to scope the set of roles

for validation. One or more interfaces with the results of the role validation can then be determined and presented to the user through the role validation tool of user interface **1558** (or interface **1568**). Such interfaces may include an assessment metric (also refer to as a score) for the set of these roles, where the metric is a reflection, for example, of the quality or health of the structure of the set of roles. By assessing the health of the roles structure, such metrics may be useful for compliance purposes or to assist in optimizing the role structure or more generally streamlining role management for the enterprise.

To provide such a role validation or assessment tool, identity management system **1560** may include role assessor **1590**. Role assessor **1590** may include an interface **1592**. Interfaces **1568** of the identity management system **1560** or interface **1558** may present one or more interfaces which may be used to access risk assessment, including the validation of roles based on an identity graph in the graph data store **1566** or subgraphs thereof. When a request for role assessment is received from the user through the user interface **1558** (or interface **1568**), a request to perform role assessment can be submitted to the role assessor **1590** from the identity management system **1550** (e.g., or user interface **1558** or other component of identity management system **1550**) through the interface **1592**, where the request may include, for example, an identification of a type of role graph to use (e.g., an access role graph or a concurrency role graph) and zero or more other criteria or attributes to utilize when determining a graph or subgraph to utilize when performing role assessment. These criteria may include, for example, zero or more scoping attributes that may have been provided by the user through the user interface or zero or more thresholds (e.g., a pruning threshold or concurrency count (support) threshold or the like) that may have been provided by the user through the user interface. Note here, that while the identity management system **1550** has been depicted in the diagram as separate and distinct from the identity management system **1560** and coupled to identity management system **1560** over a computer network **1504**, it will be realized that such an identity management system **1550** and identity management system **1560** may be deployed as part of the same identity management system or different identity management system, as a cloud based application or set of services, or in another configuration entirely.

As such, when a request for role validation and the zero or more associate scoping (or other) attributes are received through the role assessor interface **1592**, the role assessor **1590** may preform role assessment and generate a health metric as discussed. In particular, in one embodiment, the role assessor **1590** may determine a role graph to utilize for role validation based on the identification of the type of role graph to utilize received in the request along with the zero or more scoping attributes received in the request. Specifically, the role assessor **1590** may query or otherwise access the graph data store **1566** to obtain the specified type of role graph

As an example, if an access role graph is specified in the request the role assessor **1590** may query the graph data store **1566** to obtain the access role graph comprising roles nodes and access similarity relationships between those role nodes (e.g., and which may include the entitlement nodes or identity nodes and edges associated with those role nodes or access similarity relationships). If there are any scoping attributes, the set of role nodes of the obtained access role graph may be further scoped by those scoping attributes such that the access role graph only includes role nodes having (or not having) such attributes). For example, it may be desired to scope the role nodes for only roles associated with certain locations or departments such that the access role graph for which role assessment is performed only includes role nodes from those locations or departments.

Specifically, in one embodiment a scoped access role (sub)graph associated with a scoping attribute may be determined from an identity graph by querying the identity graph in the graph data store **1566** based on the scoping attribute to find the role (or other) nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the role nodes of the graph associated with the scoping attribute and determining the entitlement and identity nodes and edges along any path of the identity graph originating with each of those scoped role nodes.

Similarly, if a concurrency role graph is specified in the request the role assessor **1590** may query the graph data store **1566** to obtain the concurrency role graph comprising roles nodes and concurrent similarity relationships between those role nodes (e.g., and which may include the entitlement nodes or identity nodes and edges associated with those role nodes or concurrent similarity relationships). If there are any scoping attributes, the set of role nodes of the obtained concurrent role graph may be further scoped by those scoping attributes such that the access role graph only includes role nodes having (or not having) such attributes).

It will be realized, that in some embodiments, even if one type of role graph or the other has been specified in a request received from the user it may be useful to obtain both types of role graphs (e.g., both access and concurrent role graphs) at the time the graph data store **1566** is queried, as both types of role graphs may be utilized in some embodiments to compute a health metric for the set of (e.g., scoped roles). Moreover, it may also be a relatively similar query to obtain both types of role graphs in the same query as each type of role graph may include the same or similar role nodes and may differ only in a type of relationship (e.g., access similarity or concurrent similarity) between those role nodes.

Once the role graph to utilize for role validation has been obtained (e.g., concurrent role graph, access role graph or both), the obtained role graph may be pruned according to any pruning threshold received. This pruning may remove any edges between roles associated with the requested role graph type whose similarity weight may fall below the pruning threshold. Specifically, in certain embodiments, if an access role graph has been specified in the request all access similarity relationships between role nodes whose similarity weight falls below the pruning threshold may be removed from the obtained role graph during pruning, while if a concurrency role graph has been specified all concurrent similarity relationships whose similarity weight falls below the pruning threshold may be removed from obtained role graph during pruning. Thus, by pruning an access role graph all access similarity relationship edges whose weight falls below the pruning threshold may be removed from the role graph while by pruning a concurrency role graph any concurrent similarity relationship edges whose weight falls below the pruning threshold may be removed from the role graph.

Additionally, when pruning a concurrency role graph and concurrent similarity relationship edges whose support falls below any give support threshold (also known as a concurrent or concurrency threshold) may also be removed. As discussed, this support may be defined as the intersection of identities that are shared between two roles and thus that is

used to define the weight of a concurrent similarity relationship between two role nodes. However, highly concurrent roles may only share a few identities, and thus are particularly representative of particular (e.g., risky) significance. Thus, by filtering out (e.g., highly) concurrent roles that share only a few identities, the resulting concurrency role graph may be made more meaningful or representative of roles of interest.

Once the role graph to utilize has been obtained, it can be presented to the user through the user interface **1558** or interface **1568** in response to the originally received request. Additionally, one or more health metrics reflective of the quality of the role structure of access models of the enterprise (or the portions of the enterprise represented in the obtained role graph) may be determined and presented to the user as well. One metric may be related to a population coverage. Specifically, a metric may be determined based on the relative or absolute number of identities associated with each role (or roles not originally assigned to the identity, or identities not assigned to the role, at the time of creation of the role or identity). Roles with too large of a coverage of identities may indicate poor access management.

In some embodiments, metrics used in graph theory may provide a starting point to standardize quality scoring for these role structures and access models. In one embodiment, when generating a score based on an access role graph, a difference between a theoretically best graph structure for role structure within an enterprise may be used as a baseline to generate such metrics. For example, it may be understood that the more cliques in an access role graph, the harder it may be for an enterprise to manage its role structure.

To illustrate, brief reference is made to FIG. **16**, wherein a set of example access role graphs are presented. It will be noted that a theoretically worse access role graph for a role structure may be thought of as a clique as represented in graph **1604**, where every role is related (e.g., by an access similarity relationship) to every other role such that in a role graph representing such a structure there would be order n^2 edges in the graph **1604**. Conversely, a theoretically best access role graph for a role structure may be thought of as a set of unrelated role nodes (e.g., all singletons) as represented in graph **1602**, where every role is unrelated (e.g., by an access similarity relationship) to every other role such that in an access role graph representing such a structure there would be 0 edges in the graph **1602**.

Returning to FIG. **15**, thus, a score may be determined using an access role graph based on the number of edges in the determine access role graph relative to either 0 edges or n^2 edges. In this manner, the metric determined would be reflective of the access role structure in relation to a theoretically perfect (or worse) access role structure. Such a metric may be referred to as an access (density) ratio.

This access (density) ratio may be a scaling metric that measures the deviation from an optimal structure of an access role graphs; ranging between the worst-case access "containing high magnitude of clique (pseudo-clique) structures" (e.g. clique with similarity>70%) to the best-case access "approaching an edgeless (graph) structure" (e.g., how far off of achieving optimal structure above certain pruning threshold, like 70%). This access ratio may thus be determined based on an order (e.g., O) of the number of edges vs number of (role) nodes scaling along O(n) vs $O(n^2)$ across entire access role graph

Similarly, for a concurrency role graph, a concurrency (density) ratio may be determined as a health metric. Such a concurrency (density) ratio may be a scaling that measures the deviation from optimal structure of concurrency graphs taking into account the support (intersection of identities); ranging between worst-case (e.g. a clique or a dense enough graph) to best-case "edgeless singletons." The determination of such a metric may be substantially similar to the determination of an access ratio as discussed. However, in some cases to determine such a concurrency ratio the edges (e.g., concurrent similarity edges) may be filtered or weighted by a support threshold (e.g., both in cases where they are, or are not, pruned in the determination of the concurrency graph).

Other metrics, or combinations or metrics, may be determined without loss of generality. For example, a combination of an access ratio and concurrency ratio may be determined where these ratios may be weighted, balanced, mixed, averaged or combined in some other manner to determine one or more scores reflecting overall well-being (health) of access governance. Thus, a determined metric may be presented in association with a role graph (an access role graph or concurrency role graph, or both) through the user interface **1558** or interface **1568** in response to the originally received request, where the presented metric may have been determined based on the presented role graph.

Users associated with an enterprise could thus utilize such a role validation or access awareness interface to evaluate and validate their existing role structure to explore hierarchical relationships between existing roles; profile, re-provision, or label (e.g., tag) highly similar existing roles, consolidate and label existing roles that are heavily concurrent within certain populations, or evaluate the health of an entire (or portion of) a role structure based on the scoring system or visual depiction of a role graph.

Similarly, users involved in active access modeling or governance process (e.g., using role mining capabilities) could utilize the role validation or access awareness interface for decisions related to prioritizing roles based on the novelty with respect to existing roles, provisioning newly discovered roles with significantly high contrast to existing roles, merging, profiling, or labeling highly similar existing roles, enhancing access interpretability and enabling detection of potential risk based on security policies or, validating the impact of provisioned roles under a current role structure.

In one embodiment, such an interface may allow a user to effectively evaluate the effect of the removal of one or more roles from, or addition of one or more roles to, their existing role structure. Specifically, a role validation tool presented through the user interface **1558** or interface **1568** may allow a user to add (or remove) a specific role from a set of roles (e.g., roles associated with a presented role graph and health metric). The role assessor **1590** can then determine or update the role graph (e.g., the access role graph or concurrency role graph, or both) based on the addition (or removal) of this role and determine an update health metric based on the updated role graph. This updated graph and updated role health metric may be presented to the user through the interface **1558** or interface **1568**. Such an updated role graph and health metric may be presented, for example, alongside the originally presented role graph and health metric so a user may ascertain the effect of the addition (or removal) of that role both visually through the difference in the graph structures presented, and quantitatively through the difference in the health scores presented.

Moreover, such an interface may allow present a user with recommendations (e.g., a risk amelioration recommendation) regarding modifications to an existing role structure. For example, role assessor **1590** may perform clustering on a determined role graph (e.g., a concurrency role graph) to cluster the role nodes of such a role graph. Such clustering

may be accomplished as discussed elsewhere herein. Cliques or pseudo-cliques of roles determined from such clustering may then be determined and presented to users through the interface **1558** or **1568** for the user to consider consolidation of such roles. In the case of clustering of a concurrency role graph, it may be recommended to define a portfolio role that includes all the roles of an identified clique or pseudo-clique such that the portfolio role may be assigned to the identities that have been granted the roles comprising that clique or pseudo-clique. In that way, roles that share a similar group of identities (e.g., where the same set of identities share a set of roles) may be bundled together and an overarching role (referred to as a portfolio role) may be defined such that the bundle of similar roles may be manage as a group using the portfolio role. Thus, using embodiments, the actual scope of identities (e.g., a user population) for which roles can be consolidated to reduce use of resources in role management for that specific population and defining or assigning roles for that population. More generally, then, by reducing the number of roles or the interactions with these roles, the number of both computing resources and man hours required for such identity governance may be reduced, along with the commensurate cost to the enterprise of such identity management.

It may be helpful to an understanding of embodiments to briefly discuss an example role graph. Looking then at FIG. **17**, a graphical depiction of an identity graph **1700** (or portion thereof) that includes an example role graph (or portion thereof) is depicted. Here, nodes are represented by circles and relationships are represented by the directional arrows between the nodes. Such a role graph **1700** may represent roles, identities or entitlements, their association, and the degree of access similarity or concurrent similarity (or both) between roles represented by the role nodes. Thus, for example, role nodes **1708***a*, **1708***b* have the label "Role" indicating they are role nodes. Role nodes **1708***a*, **1708***b* are associated with a set of properties that define the attributes or data of that role node **1708***a*, **1708***b*, including here that the "id" of role node **1708***a* is "Role_**4562**" and the "id" of role node **1708***b* is "Role **3128**".

Similarly identity nodes **1702***a*, **1702***b*, **1702***c* have the label "Identity" indicating they are identity nodes. Identity nodes **1702***a*, **1702***b*, **1702***c* are associated with a set of properties that define the attributes or data of that identity node. For example, identity node **1702***a* is shown as being associated with a set of properties that define the attributes or data of that identity node **1702***a*, including here that the "id" of identity node **1702***a* is "a123", the "company" of identity node **1702***a* is "Ajax", the "dept" of identity node **1702***a* is "Engineering", the "title" of identity node **1702***a* is "Developer, and the "location" of identity node **1702***a* is "Austin".

Entitlement nodes **1704***a*, **1704***b*, **1704***c*, **1704***d* have the label "Entitlement" indicating that they are entitlement nodes. Entitlement nodes **1704***a*, **1704***b*, **1704***c*, **1704***d* are associated with a set of properties that define the attributes or data of that entitlement node. For example, entitlement node **1704***b* is shown as being associated with a set of properties that define the attributes or data of that entitlement node **1704***b*, including here that the "id" of entitlement node **1704***b* is "ad179", and the "source" of entitlement node **1704***b* is "Active Directory". Entitlement node **1704***a* is shown as being associated with a set of properties that define the attributes or data of that entitlement node **1704***a*, including here that the "id" of entitlement node **1704***a* is "ok143", and the "source" of entitlement node **1704***a* is "Okta".

Identity nodes **1702** and role nodes **1708** of the identity graph can be joined by edges formed by directed relationships **1716**. Directed relationships **1716** may represent that the identity of identity node **1702** has (represented by the labeled "HAS_ROLE" relationships **1716**) the role represented by the role nodes **1708**. For example, HAS_ROLE relationship **1716***a* represents that the identity represented by identity node **1702***a* has been assigned the role represented by role node **1708***a*. Similarly, HAS_ROLE relationship **1716***b* represents that the identity represented by identity node **1702***b* has been assigned the role represented by role node **1708***a*, HAS_ROLE relationship **1716***c* represents that the identity represented by identity node **1702***b* has been assigned the role represented by role node **1708***b*, and HAS_ROLE relationship **1716***d* represents that the identity represented by identity node **1702***c* has been assigned the role represented by role node **1708***b*.

Entitlement nodes **1704** and role nodes **1708** of the identity graph can be joined by edges formed by directed relationships **1714**. Directed relationships **1714** may represent that the role of a role node **1708** includes (represented by the labeled "HAS_ENT" relationships **1714**) the entitlement of the related entitlement node **1704**. For example, HAS_ENT relationship **1714***a* represents that the role represented by role node **1708***a* includes the entitlement represented by entitlement node **1704***a*. Similarly, HAS_ENT relationship **1714***b* represents that the role represented by role node **1708***a* includes the entitlement represented by entitlement node **1704***b*, HAS_ENT relationship **1714***c* represents that the role represented by role node **1708***b* includes the entitlement represented by entitlement node **1704***b*, HAS_ENT relationship **1714***d* represents that the role represented by role node **1708***b* includes the entitlement represented by entitlement node **1704***c* and HAS_ENT relationship **1714***e* represents that the role represented by role node **1708***b* includes the entitlement represented by entitlement node **1704***d*.

The role nodes **1708** of the identity graph may be joined by edges formed by concurrent similarity relationships **1722**. Concurrent similarity relationships **1722** may represent that the role of one role node **1708** is similar to (represented by the labeled "CONCURRENT_SIM" relationship **1722**) the role of the related role node **1708** based on shared identities which have that role. A weight may be computed for the concurrent similarity relationship **1722** the number of identities nodes **1702** which share those roles (e.g., which have HAS_ROLE relationships **1716** with both roles nodes **1708**) and the number of identities that have those roles (e.g., the number of identity nodes **1702** that have HAS_ROLE relationships **1716** with either of the roles nodes **1708**). In one embodiment, concurrent similarity relationship **1722** between role nodes **1708***a*, **1708***b* may be determined as the Intersection(number of identities nodes **1702** having roles **1708**)/Union(number of identities nodes **1702** having either of the roles **1708**). For example, here, CONCURRENT_SIM relationship **1722** may have a weight of 0.33 assigned to it.

The role nodes **1708** of the identity graph may also be joined by edges formed by access similarity relationships **1724**. Access similarity relationships **1724** may represent that the role of one role node **1708** is similar to (represented by the labeled "ACCESS_SIM" relationship **1724**) the role of the related role node **1708** based on entitlements that those roles share. A weight may be computed for the access similarity relationship **1724** based on the number of entitlement nodes **1704** shared by those roles **1708** (e.g., which have a HAS_ENT relationships **1714** with both roles nodes

1708) and the number of entitlements that those roles have (e.g., the number of entailment nodes **1704** with which either of those role nodes **1708** has a HAS_ENT relationships **1714**). In one embodiment, access similarity relationship **1724** between role nodes **1708***a*, **1708***b* may be determined as the Intersection(number of entitlement nodes **1704** having relationships with both roles nodes **1708**)/Union(number of entitlement nodes **1704** having relationships with either roles node **1708**). For example, here, ACCESS_SIM relationship **1724** may have a weight of 0.25 assigned to it. Note that both these types of similarity relationships **1722**, **1724** may be a single bidirectional relationship assigned a single similarity weight or may be bidirectional relationships that may be weighted differently based on different criteria.

As can be seen then, an identity graph may include a role graph that includes both an access role graph modeled in terms of entitlement (e.g., access) similarities between roles and a concurrency graph that is modeled in terms of identities shared between roles. In the access role graph, certain edges (e.g., ACCESS_SIM relationships **1724**) represent an access similarity relationship between two roles (e.g., nodes representing the roles) joined by that edge of the graph, where the access similarity relationship may have a weight based on the entitlements shared between the roles and the number of entitlements the roles include. In the concurrency role graph, the edges (e.g., CONCURRENT_SIM relationships **1722**) represent a concurrent similarity modeled in terms of shared identities shared between the roles. A weight may be computed for the concurrent similarity relationship based on the number of identities which share those roles and the number of identities that have those roles.

With examples of such an access role graph or concurrency role graph in mind, reference is now made to FIG. **18** where a flow diagram of one embodiment of a method for performing role assessment is depicted. Embodiments of such a method may be performed, for example by an identity management system or a role assessor of such an identity management system. Initially, a request to perform role assessment may be received (STEP **1810**). The request may include, for example, an identification of a type of role graph to use (e.g., an access role graph or a concurrency role graph) and zero or more other criteria or attributes to utilize when determining a graph or subgraph to utilize when performing role assessment. These criteria may include, for example, zero or more scoping attributes that may have been provided by the user through the user interface or zero or more thresholds (e.g., a pruning threshold or count (support) threshold or the like) that may have been provided by the user through the user interface.

In particular, in one embodiment, a role graph to utilize for role validation may be determined based on the identification of the type of role graph to utilize received in the request along with the zero or more scoping attributes received in the request (STEP **1820**). Specifically, the graph may be queried to obtain the specified type of role graph

As an example, if an access role graph is specified in the request the graph may be queried to obtain the access role graph comprising roles nodes and access similarity relationships between those role nodes (e.g., and which may include the entitlement nodes or identity nodes and edges associated with those role nodes or access similarity relationships). If there are any scoping attributes, the set of role nodes of the obtained access role graph may be further scoped by those scoping attributes such that the access role graph only includes role nodes having (or not having) such attributes). For example, it may be desired to scope the role nodes for only roles associated with certain locations or departments such that the access role graph for which role assessment is performed only includes role nodes from those locations or departments.

Specifically, in one embodiment a scoped access role (sub)graph associated with a scoping attribute may be determined from an identity graph by querying the identity graph based on the scoping attribute to find the role (or other) nodes and edges associated with the scoping attribute. Such querying may involve, for example, querying the identity graph to determine the role nodes of the graph associated with the scoping attribute and determining the entitlement and identity nodes and edges along any path of the identity graph originating with each of those scoped role nodes.

Similarly, if a concurrency role graph is specified in the request the graph may be queried to obtain the concurrency role graph comprising roles nodes and concurrent similarity relationships between those role nodes (e.g., and which may include the entitlement nodes or identity nodes and edges associated with those role nodes or concurrent similarity relationships). If there are any scoping attributes, the set of role nodes of the obtained concurrent role graph may be further scoped by those scoping attributes such that the concurrent role graph only includes role nodes having (or not having) such attributes). In some embodiments, even if one type of role graph or the other has been specified in a request received from the user it may be useful to obtain both types of role graphs (e.g., both access and concurrent role graphs) at the time the graph is queried, as both types of role graphs may be utilized in some embodiments to compute a health metric for the set of (e.g., scoped roles).

Once the role graph to utilize for role validation has been obtained (e.g., concurrent role graph, access role graph or both), the obtained role graph may be pruned according to any pruning threshold received (STEP **1830**). This pruning may remove any edges associated with the requested role graph type whose similarity weight may fall below the pruning threshold. Specifically, in certain embodiments, if an access role graph has been specified in the request all access similarity relationships between role nodes whose similarity weight falls below the pruning threshold may be removed from the obtained role graph during pruning, while if a concurrency role graph has been specified all concurrent similarity relationships whose similarity weight falls below the pruning threshold may be removed from obtained role graph during pruning.

Additionally, when pruning a concurrency role graph and concurrent similarity relationship edges whose support falls below any give support threshold may also be removed. As discussed, this support may be defined as the intersection of identities that are shared between two roles and thus that is used to define the weight of a concurrent similarity relationship between two role nodes. However, that highly concurrent roles may only share a few identities, and thus are particularly representative or of particular (e.g., risky) significance. Thus, by filtering out (e.g., highly) concurrent roles that share only a few identities, the resulting concurrency role graph may be made more meaningful or representative of roles of interest.

Once the role graph to utilize has been obtained, it can be used to present role data (e.g., including the role graph) to the user through a user interface in response to the originally received request (STEP **1850**). Additionally, in some embodiments, one or more health metrics reflective of the quality of the role structure of access models of the enterprise (or the portions of the enterprise redefected in the obtained role graph) may be determined (STEP **1840**) and presented to the user as part of the represented ole data

(STEP **1850**). Such health metrics, may for example, be determined on the structure of the role graph. One such health metric may be related to a population coverage. Specifically, a metric may be determined based on the relative or absolute number of identities associated with each role (or roles not originally assigned to the identity, or identities not assigned to the role, at the time of creation of the role or identity). Roles with too large of a coverage of identities may indicate poor access management.

In one embodiment, when generating a score based on an access role graph, a difference between a theoretically best graph structure for role structure within an enterprise may be used as a baseline to generate such metrics. For example, it may be understood that the more cliques in an access role graph, the harder it may be for an enterprise to manage its role structure. Thus, a score may be determined using an access role graph based on the number of edges in the determine access role graph relative to either 0 edges or n^2 edges. In this manner, the metric determined would be reflective of the access role structure in relation to a theoretically perfect (or worse) access role structure. Such a metric may be referred to as an access (density) ratio. Similarly, for a concurrency role graph, a concurrency (density) ratio may be determined as a health metric. Such a concurrency (density) ratio may be a scaling that measures the deviation from optimal structure of concurrency graphs taking into account the support (intersection of identities); ranging between worst-case (e.g. a clique or a dense enough graph) to best-case "edgeless singletons." In some cases to determine such a concurrency ratio, the edges (e.g., concurrent similarity edges) may be filtered or weighted by a support threshold (e.g., both in cases where they are, or are not, pruned in the determination of the concurrency graph).

Other metrics, or combinations or metrics, may be determined without loss of generality. For example, a combination of an access ratio and concurrency ratio may be determined where these ratios may be weighted, balanced, mixed, averaged or combined in some other manner to determine one or more scores reflecting overall well-being (health) of access governance. Thus, a determined metric may be presented in association with a role graph through the user interface in response to the originally received request, where the presented metric may have been determined based on the presented role graph (STEP **1850**).

Users associated with an could thus utilize such a role validation or access awareness interface to evaluate and validate their existing role structure to explore hierarchical relationships between existing roles; profile, re-provision, or label (e.g., tag) highly similar existing roles, consolidate and label existing roles that are heavily concurrent within certain populations, or evaluate the health of an entire (or portion of) a role structure based on the scoring system or visual depiction of a role graph.

Similarly, users involved in active access modeling or governance process (e.g., using role mining capabilities) could utilize the role validation or access awareness interface for decisions related to prioritizing roles based on the novelty with respect to existing roles, provisioning newly discovered roles with significantly high contrast to existing roles, merging, profiling, or labeling highly similar existing roles, enhancing access interpretability and enabling detection of potential risk based on security policies or validating the impact of provisioned roles under a current role structure.

In one embodiment, such an interface may allow a user to effectively evaluate the effect of the removal of one or more roles from, or addition of one or more roles to, their existing role structure. Specifically, a role validation interface presented through the user interface may allow a user to add (or remove) a specific role from a set of roles (e.g., roles associated with a presented role graph and health metric). The role graph (e.g., the access role graph or concurrency role graph, or both) may be updated based on the addition (or removal) of this role and determine an updated health metric based on the updated role graph. This updated graph and updated role health metric may be presented to the user through the interface. Such an updated role graph and health metric may be presented, for example, alongside the originally presented role graph and health metric so a user may ascertain the effect of the addition (or removal) of that role both visually through the difference in the graph structures presented, and quantitatively through the difference in the health scores presented.

Moreover, such an interface may allow present a user with recommendations regarding modifications to an existing role structure. For example, it may be recommended to define a portfolio role that includes a set of roles within a presented graph such that the portfolio role may be assigned to the identities that have been granted the roles comprising that set of roles. In that way, roles that share a similar group of identities (e.g., where the same set of identities share a set of roles) may be bundled together and an overarching role (referred to as a portfolio role) may be defined such that the bundle of similar roles may be managed as a group using the portfolio role. Thus, using embodiments, the actual scope of identities (e.g., a user population) for which roles can be consolidated to reduce use of resources in role management for that specific population and defining or assigning roles for that population. More generally, then, by reducing the number of roles or the interactions with these roles, the number of both computing resources and man hours required for such identity governance may be reduced, along with the commensurate cost to the enterprise of such identity management.

Certain example interfaces that may be used for roles assessment in embodiments of an identity management system are depicted in FIGS. **19**A, **19**B, **19**C and **19**D. Looking first at FIG. **19**A, one embodiment of an interface for role assessment is presented. The user may utilize interface **1900** to role graph selection portion **1902** of the interface to select whether a concurrency role graph or an access role graph is desired (here two tabs associated with each type of role graph). In the example depicted, the user has selected an access role graph for role assessment. The user may also be presented with a threshold selection portion **1904** of the interface (in this case a slider bar), where the user may select a pruning threshold to utilize for pruning the (e.g., similarity relationships) of the role graph to be determined and presented in the interface **1900**. In this case, as the role graph will be an access role graph, the threshold selection portion presents a threshold selection portion **1904** for an access similarity relationship similarity threshold. An access role graph generated based on the user's selection may then be presented in the interface **1900** where the points presented represent the role nodes of the graph and the edges represent the access similarity relationships between those roles (e.g., where the weights on those access similarity relationships are all at or above the selected pruning threshold). Other visual indicators may also be used with respect to the presented graph. For example, the size of a point for a role node may reflect the number of identities having that role. FIG. **19**B depicts another embodiment of an interface for role assessment where an access role graph is being presented.

Turning now to FIG. **19**C, one embodiment of an interface for role assessment is presented where the user has utilized role graph selection portion **1902** interface **1900** to select that a concurrency role graph is desired. Now, threshold selection portion **1904** of the interface **1900** may present selection mechanisms for a concurrency similarity relationship similarity threshold and a concurrent count (support) threshold. A concurrency role graph generated based on the user's selection may then be presented in the interface **1900** where the points presented represent the role nodes of the graph and the edges represent the concurrency similarity relationships between those roles (e.g., where the weights on those concurrency similarity relationships are all at or above the selected pruning threshold and have at least the selected concurrent count). Other visual indicators may also be used with respect to the presented graph. For example, the thickness of an edge may reflect the number of identities shared between the two roles (e.g., thicker lines indicate more identities shared between the roles). FIG. **19**D depicts another embodiment of an interface for role assessment where a concurrency role graph is being presented.

Those skilled in the relevant art will appreciate that the invention can be implemented or practiced with other computer system configurations including, without limitation, multi-processor systems, network devices, mini-computers, mainframe computers, data processors, and the like. Embodiments can be employed in distributed computing environments, where tasks or modules are performed by remote processing devices, which are linked through a communications network such as a LAN, WAN, and/or the Internet. In a distributed computing environment, program modules or subroutines may be located in both local and remote memory storage devices. These program modules or subroutines may, for example, be stored or distributed on computer-readable media, including magnetic and optically readable and removable computer discs, stored as firmware in chips, as well as distributed electronically over the Internet or over other networks (including wireless networks). Example chips may include Electrically Erasable Programmable Read-Only Memory (EEPROM) chips. Embodiments discussed herein can be implemented in suitable instructions that may reside on a non-transitory computer readable medium, hardware circuitry or the like, or any combination and that may be translatable by one or more server machines. Examples of a non-transitory computer readable medium are provided below in this disclosure.

Although the invention has been described with respect to specific embodiments thereof, these embodiments are merely illustrative, and not restrictive of the invention. Rather, the description is intended to describe illustrative embodiments, features and functions in order to provide a person of ordinary skill in the art context to understand the invention without limiting the invention to any particularly described embodiment, feature or function, including any such embodiment feature or function described. While specific embodiments of, and examples for, the invention are described herein for illustrative purposes only, various equivalent modifications are possible within the spirit and scope of the invention, as those skilled in the relevant art will recognize and appreciate.

As indicated, these modifications may be made to the invention in light of the foregoing description of illustrated embodiments of the invention and are to be included within the spirit and scope of the invention. Thus, while the invention has been described herein with reference to particular embodiments thereof, a latitude of modification, various changes and substitutions are intended in the foregoing disclosures, and it will be appreciated that in some instances some features of embodiments of the invention will be employed without a corresponding use of other features without departing from the scope and spirit of the invention as set forth. Therefore, many modifications may be made to adapt a particular situation or material to the essential scope and spirit of the invention.

Reference throughout this specification to "one embodiment", "an embodiment", or "a specific embodiment" or similar terminology means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment and may not necessarily be present in all embodiments. Thus, respective appearances of the phrases "in one embodiment", "in an embodiment", or "in a specific embodiment" or similar terminology in various places throughout this specification are not necessarily referring to the same embodiment. Furthermore, the particular features, structures, or characteristics of any particular embodiment may be combined in any suitable manner with one or more other embodiments. It is to be understood that other variations and modifications of the embodiments described and illustrated herein are possible in light of the teachings herein and are to be considered as part of the spirit and scope of the invention.

In the description herein, numerous specific details are provided, such as examples of components and/or methods, to provide a thorough understanding of embodiments of the invention. One skilled in the relevant art will recognize, however, that an embodiment may be able to be practiced without one or more of the specific details, or with other apparatus, systems, assemblies, methods, components, materials, parts, and/or the like. In other instances, well-known structures, components, systems, materials, or operations are not specifically shown or described in detail to avoid obscuring aspects of embodiments of the invention. While the invention may be illustrated by using a particular embodiment, this is not and does not limit the invention to any particular embodiment and a person of ordinary skill in the art will recognize that additional embodiments are readily understandable and are a part of this invention.

Embodiments discussed herein can be implemented in a set of distributed computers communicatively coupled to a network (for example, the Internet). Any suitable programming language can be used to implement the routines, methods or programs of embodiments of the invention described herein, including R, Python, C, C++, Java, JavaScript, HTML, or any other programming or scripting code, etc. Other software/hardware/network architectures may be used. Communications between computers implementing embodiments can be accomplished using any electronic, optical, radio frequency signals, or other suitable methods and tools of communication in compliance with known network protocols.

Although the steps, operations, or computations may be presented in a specific order, this order may be changed in different embodiments. In some embodiments, to the extent multiple steps are shown as sequential in this specification, some combination of such steps in alternative embodiments may be performed at the same time. The sequence of operations described herein can be interrupted, suspended, or otherwise controlled by another process, such as an operating system, kernel, etc. The routines can operate in an operating system environment or as stand-alone routines. Functions, routines, methods, steps and operations described herein can be performed in hardware, software, firmware or any combination thereof.

Embodiments described herein can be implemented in the form of control logic in software or hardware or a combination of both. The control logic may be stored in an information storage medium, such as a computer-readable medium, as a plurality of instructions adapted to direct an information processing device to perform a set of steps disclosed in the various embodiments. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will appreciate other ways and/or methods to implement the invention.

A "computer-readable medium" may be any medium that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, system or device. The computer readable medium can be, by way of example only but not by limitation, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, system, device, propagation medium, or computer memory. Such computer-readable medium shall generally be machine readable and include software programming or code that can be human readable (e.g., source code) or machine readable (e.g., object code). Examples of non-transitory computer-readable media can include random access memories, read-only memories, hard drives, data cartridges, magnetic tapes, floppy diskettes, flash memory drives, optical data storage devices, compact-disc read-only memories, and other appropriate computer memories and data storage devices.

As used herein, the terms "comprises," "comprising," "includes," "including," "has," "having," or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a process, product, article, or apparatus that comprises a list of elements is not necessarily limited only those elements but may include other elements not expressly listed or inherent to such process, product, article, or apparatus.

Furthermore, the term "or" as used herein is generally intended to mean "and/or" unless otherwise indicated. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present). As used herein, a term preceded by "a" or "an" (and "the" when antecedent basis is "a" or "an") includes both singular and plural of such term, unless clearly indicated within the claim otherwise (i.e., that the reference "a" or "an" clearly indicates only the singular or only the plural). Also, as used in the description herein and throughout the meaning of "in" includes "in" and "on" unless the context clearly dictates otherwise.

What is claimed is:

1. An identity management system, comprising:

a data store;

a processor;

a non-transitory, computer-readable storage medium, including computer instructions, when executed by the processor, cause the system to perform the steps of:

obtaining identity management data from one or more identity management systems in a distributed enterprise computing environment, the identity management data comprising data on a set of roles, a set of entitlements, and a set of identities, wherein the set of roles, set of entitlements and set of identities are utilized in identity management in the distributed enterprise computing environment;

evaluating the identity management data to determine the set or roles, identities of the set of identities associated with each of the set of roles and entitlements of the set of entitlements associated with the set of roles;

generating a first role graph from the identity management data by:

creating a node of the first role graph for each of the determined set of roles;

for each first identity and second identity that share at least one entitlement of the set of entitlements or at least one identity of the set of identities, creating an edge of the first role graph between a first node representing a first role and a second node of the first role graph representing a second role; and

generating a weight for each edge of the first role graph between each first node representing the first role and second node representing the second role based on the at least one entitlement or the at least one identity shared between the first role represented by the first node and the second role represented by the second node;

storing the first role graph in the data store;

obtaining a pruning threshold associated with the first role graph;

pruning the set of edges of the first role graph based on the pruning threshold and the weight for each of the set of edges to generate a second role graph;

storing the second role graph in the data store;

determining a health metric for the set of roles associated with the distributed enterprise computing environment based on a structure of the second role graph; and

presenting the health metric and the second role graph to a user through a user interface.

2. The system of claim **1**, wherein the each edge is an access similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is an access similarity weight based on a number of shared entitlements between the first role represented by the first node and the second role represented by the second node.

3. The system of claim **1**, wherein the each edge is a concurrency similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is a concurrent similarity weight based on a number of identities that have both the first role represented by the first node and the second role represented by the second node.

4. The system of claim **3**, wherein the pruning of the first role graph is based on a concurrency count comprising the number of identities that include both roles.

5. The system of claim **1**, wherein the health metric for the set of roles associated with the distributed enterprise computing environment is a scaling metric that measures a deviation from an optimal structure of the second role graph.

6. The system of claim **5**, wherein the deviation from the optimal structure is determined based on the number of edges in the second role graph.

7. The system of claim **1**, wherein the health metric is based on a number of identities or entitlements associated with each role represented in the second role graph.

8. A method for managing roles, comprising the steps of:

obtaining identity management data from one or more identity management systems in a distributed enterprise computing environment via one or more of computer network connections, the identity management data comprising data on a set of roles, a set of entitlements, and a set of identities, wherein the set of roles, set of entitlements and set of identities are utilized in identity management in the distributed enterprise computing environment;

evaluating the identity management data to determine the set or roles, identities of the set of identities associated with each of the set of roles and entitlements of the set of entitlements associated with the set of roles;

generating a first role graph from the identity management data by:

creating a node of the first role graph for each of the determined set of roles;

for each first identity and second identity that share at least one entitlement of the set of entitlements or at least one identity of the set of identities, creating an edge of the first role graph between a first node representing a first role and a second node of the first role graph representing a second role; and

generating a weight for each edge of the first role graph between each first node representing the first role and second node representing the second role based on the at least one entitlement or the at least one identity shared between the first role represented by the first node and the second role represented by the second node;

storing the first role graph in the data store;

obtaining a pruning threshold associated with the first role graph;

pruning the set of edges of the first role graph based on the pruning threshold and the weight for each of the set of edges to generate a second role graph;

storing the second role graph;

determining a health metric for the set of roles associated with the distributed enterprise computing environment based on a structure of the second role graph; and

presenting the health metric and the second role graph to a user through a user interface.

9. The method of claim **8**, wherein the each edge is an access similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is an access similarity weight based on a number of shared entitlements between the first role represented by the first node and the second role represented by the second node.

10. The method of claim **8**, wherein the each edge is a concurrency similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is a concurrent similarity weight based on a number of identities that have both the first role represented by the first node and the second role represented by the second node.

11. The method of claim **10**, wherein the pruning of the first role graph is based on a concurrency count comprising the number of identities that include both roles.

12. The method of claim **8**, wherein the health metric for the set of roles associated with the distributed enterprise computing environment is a scaling metric that measures a deviation from an optimal structure of the second role graph.

13. The method of claim **12**, wherein the deviation from the optimal structure is determined based on the number of edges in the second role graph.

14. The method of claim **8**, wherein the health metric is based on a number of identities or entitlements associated with each role represented in the second role graph.

15. A non-transitory computer readable medium, comprising computer instructions, when executed by a computer processor, cause the computer process to perform the steps of:

obtaining identity management data from one or more identity management systems in a distributed enterprise computing environment, the identity management data comprising data on a set of roles, a set of entitlements, and a set of identities, wherein the set of roles, set of entitlements and set of identities are utilized in identity management in the distributed enterprise computing environment;

evaluating the identity management data to determine the set or roles, identities of the set of identities associated with each of the set of roles and entitlements of the set of entitlements associated with the set of roles;

generating a first role graph from the identity management data by:

creating a node of the first role graph for each of the determined set of roles, for each first identity and second identity that share at least one entitlement of the set of entitlements or at least one identity of the set of identities, creating an edge of the first role graph between a first node representing a first role and a second node of the first role graph representing a second role; and

generating a weight for each edge of the first role graph between each first node representing the first role and second node representing the second role based on the at least one entitlement or the at least one identity shared between the first role represented by the first node and the second role represented by the second node;

storing the first role graph in the data store;

obtaining a pruning threshold associated with the first role graph;

pruning the set of edges of the first role graph based on the pruning threshold and the weight for each of the set of edges to generate a second role graph;

storing the second role graph;

determining a health metric for the set of roles associated with the distributed enterprise computing environment based on a structure of the second role graph; and

presenting the health metric and the second role graph to a user through a user interface.

16. The non-transitory computer readable medium of claim **15**, wherein the each edge is an access similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is an access similarity weight based on a number of shared entitlements between the first role represented by the first node and the second role represented by the second node.

17. The non-transitory computer readable medium of claim **15**, wherein the each edge is a concurrency similarity relationship and the weight generated for each edge of the first role graph between each first node and second node is a concurrent similarity weight based on a number of identities that have both the first role represented by the first node and the second role represented by the second node.

18. The non-transitory computer readable medium of claim **17**, wherein the pruning of the first role graph is based on a concurrency count comprising the number of identities that include both roles.

19. The non-transitory computer readable medium of claim **15**, wherein the health metric for the set of roles associated with the distributed enterprise computing environment is a scaling metric that measures a deviation from an optimal structure of the second role graph.

20. The non-transitory computer readable medium of claim **19**, wherein the deviation from the optimal structure is determined based on the number of edges in the second role graph.

21. The non-transitory computer readable medium of claim **15**, wherein the health metric is based on a number of

identities or entitlements associated with each role represented in the second role graph.

* * * * *